Central limit theorems and confidence sets in the calibration of Lévy models and in deconvolution

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Dr. Rer. Nat.
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
Humboldt-Universität zu Berlin

von
Dipl.-Math. Jakob Söhl

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II:
Prof. Dr. Elmar Kulke

Gutachter:
1. Markus Reiß
2. Vladimir Spokoiny
3. Richard Nickl

Tag der Verteidigung: 21.03.2013

Abstract

Central limit theorems and confidence sets are studied in two different but related nonparametric inverse problems, namely in the calibration of an exponential Lévy model and in the deconvolution model.

In the first set–up, an asset is modeled by an exponential of a Lévy process, option prices are observed and the characteristic triplet of the Lévy process is estimated. We show that the estimators are almost surely well–defined. To this end, we prove an upper bound for hitting probabilities of Gaussian random fields and apply this to a Gaussian process related to the estimation method for Lévy models. We prove joint asymptotic normality for estimators of the volatility, the drift, the intensity and for pointwise estimators of the jump density. Based on these results, we construct confidence intervals and sets for the estimators. We show that the confidence intervals perform well in simulations and apply them to option data of the German DAX index.

In the deconvolution model, we observe independent, identically distributed random variables with additive errors and we estimate linear functionals of the density of the random variables. We consider deconvolution models with ordinary smooth errors. Then the ill–posedness of the problem is given by the polynomial decay rate with which the characteristic function of the errors decays. We prove a uniform central limit theorem for the estimators of translation classes of linear functionals, which includes the estimation of the distribution function as a special case. Our results hold in situations, for which a $\sqrt{n}$–rate can be obtained, more precisely, if the L^2–Sobolev smoothness of the functionals is larger than the ill–posedness of the problem.

Zusammenfassung

Zentrale Grenzwertsätze und Konfidenzmengen werden in zwei verschiedenen, nichtparametrischen, inversen Problemen ähnlicher Struktur untersucht, und zwar in der Kalibrierung eines exponentiellen Lévy–Modells und im Dekonvolutionsmodell.

Im ersten Modell wird eine Geldanlage durch einen exponentiellen Lévy–Prozess dargestellt, Optionspreise werden beobachtet und das charakteristische Tripel des Lévy–Prozesses wird geschätzt. Wir zeigen, dass die Schätzer fast sicher wohldefiniert sind. Zu diesem Zweck beweisen wir eine obere Schranke für Trefferwahrscheinlichkeiten von gaußschen Zufallsfeldern und wenden diese auf einen Gauß–Prozess aus der Schätzmethode für Lévy–Modelle an. Wir beweisen gemeinsame asymptotische Normalität für die Schätzer von Volatilität, Drift und Intensität und für die punktweisen Schätzer der Sprungdichte. Basierend auf diesen Ergebnissen konstruieren wir Konfidenzintervalle und –mengen für die Schätzer. Wir zeigen, dass sich die Konfidenzintervalle in Simulationen gut verhalten, und wenden sie auf Optionsdaten des DAX an.

Im Dekonvolutionsmodell beobachten wir unabhängige, identisch verteilte Zufallsvariablen mit additiven Fehlern und schätzen lineare Funktionale der Dichte der Zufallsvariablen. Wir betrachten Dekonvolutionsmodelle mit gewöhnlich glatten Fehlern. Bei diesen ist die Schlechtgestelltheit des Problems durch die polynomielle Abfallrate der charakteristischen Funktion der Fehler gegeben. Wir beweisen einen gleichmäßigen zentralen Grenzwertsatz für Schätzer von Translationsklassen linearer Funktionale, der die Schätzung der Verteilungsfunktion als Spezialfall enthält. Unsere Ergebnisse gelten in Situationen, in denen eine $\sqrt{n}$–Rate erreicht werden kann, genauer gesagt gelten sie, wenn die L^2–Sobolev–Glattheit der Funktionale größer als die Schlechtgestelltheit des Problems ist.

Contents

1 Introduction

Central limit theorems for estimators are of fundamental interest since they allow to assess the reliability of estimators and to construct confidence sets which cover the unknown parameters or functions with a prescribed probability. We focus on nonparametric inverse problems and study two different but related models. In the first one, an asset price is modeled by an exponential of a Lévy process. Option prices of the asset are observed and the aim is to estimate the characteristic triplet of the Lévy process. The second model is deconvolution, where independent, identically distributed random variables with additive error are observed and the aim is statistical inference on the distribution of the random variables.

In both models, in the estimation of the Lévy process and in the deconvolution, we study central limit theorems for estimators that are based on Fourier methods. In the first set–up, the price of an asset (S_t) follows under the risk–neutral measure an exponential Lévy model

$$S_t = Se^{rt+L_t} \quad \text{with a Lévy process } (L_t) \text{ for } t \geqslant 0,$$

where $S > 0$ the present value of the stock and $r \geqslant 0$ is the riskless interest rate. Based on prices of European options with the underlying (S_t), we calibrate the model by estimating the characteristic triplet of (L_t), which consists of the volatility, the drift and the Lévy measure. We construct confidence sets for the characteristic triplet. This is of particular importance since the calibrated model is the basis for pricing and hedging. The calibration problem is closely related to the classical nonparametric inverse problem of deconvolution. On the one hand the law of the continuous part is convolved with the law of the jump part of the Lévy process. On the other hand the Lévy measure is convolved with itself in the marginal distribution of the jump part. Besides being itself an interesting problem with many applications, the deconvolution model exhibits the same underlying structure as the nonlinear estimation of the characteristic triplet of a Lévy process and is easier to analyze since it is linear. In the deconvolution model, we observe n random variables

$$Y_j = X_j + \varepsilon_j, \qquad j = 1, \ldots, n,$$

where the X_j are identically distributed with density f_X, the ε_j are identically distributed with density f_ε and where $X_1, \ldots, X_n, \varepsilon_1, \ldots, \varepsilon_n$ are independent. The aim is to estimate the distribution function of the X_j or, more precisely, linear functionals $\int \zeta(x-t) f_X(x)\,\mathrm{d}x$ of the density f_X, where the special case $\zeta = \mathbf{1}_{(-\infty,0]}$ leads to the estimation of the distribution function. Since the central limit theorems show that the estimators are asymptotically normally distributed, we also speak of asymptotic normality. In both problems, we determine the joint asymptotic distribution of the estimators. In the deconvolution problem, we even prove a uniform central limit theorem meaning that the asymptotic normality holds uniformly over all $t \in \mathbb{R}$. Based on the joint asymptotic distribution of the estimators, we construct confidence intervals and joint confidence sets. The uniform convergence in the deconvolution paves the way for the construction of confidence bands.

Lévy processes are widely used in financial modeling, since they allow to reproduce many stylized facts well. Exponential Lévy models generalize the classical model by Black and Scholes (1973) by accounting in addition to volatility and drift for jumps in the price process. They are capable of modeling not only a volatility smile or skew but also the effect that the smile or skew is more pronounced for shorter maturities. For a recent review on pricing and hedging in exponential Lévy models see Tankov (2011). The calibration of exponential Lévy models has mainly focused on parametric models, cf. Barndorff-Nielsen (1998); Carr et al. (2002); Eberlein et al. (1998) and the references therein. First nonparametric calibration procedures for Lévy models were proposed by Cont and Tankov (2004b, 2006) as well as by Belomestny and Reiß (2006a). In these approaches no parametrization is assumed on the jump density and thus the model misspecification is reduced. In both methods, the calibration is based on prices of European call and put options. Cont and Tankov introduce a least squares estimator penalized by relative entropy. Belomestny and Reiß propose the spectral calibration method and show that it achieves the minimax rates of convergence. The spectral calibration method is designed for finite intensity Lévy processes with Gaussian component and it is based on a regularization by a spectral cut–off. We show asymptotic normality as well as construct confidence sets and intervals for the spectral calibration method. Similar methods were also applied by Belomestny (2010) to estimate the fractional order of regular Lévy processes of exponential type, by Belomestny and Schoenmakers (2011) to calibrate a Libor model and by Trabs (2012) to estimate self–decomposable Lévy processes.

The estimation of Lévy processes from direct observations has been studied for high–frequency and for low–frequency observations. For high–frequency observations the time between observations tends to zero as the number of observations grows, while for low–frequency observations the time between observations is fixed. As a starting point for high–frequency observations, Figueroa-López and Houdré (2006) estimate a Lévy process nonparametrically from continuous observations. Nonparametric estimation from discrete observations at high frequency is treated by Figueroa-López (2009) or by Comte and Genon-Catalot (2009, 2011). Nonparametric estimation of Lévy processes from low–frequency observations has been studied for the estimation of functionals by Neumann and Reiß (2009), for finite intensity Lévy processes with Gaussian component by Gugushvili (2009) or for pure jump Lévy processes of finite variation via model selection by Comte and Genon-Catalot (2010) and by Kappus (2012).

We prove asymptotic normality for the spectral calibration method, where the Lévy process is observed only indirectly since the method is based on option prices. The indirect observation scheme does not correspond to direct observations at high frequency but at low frequency. Our asymptotic normality results are for estimators of the volatility, the drift, the intensity and pointwise for estimators of the jump density. These theorems on asymptotic normality belong to the main results of this thesis and are also available in Söhl (2012). This calibration problem is a statistical inverse problem, which is completely different in the mildly ill–posed case of volatility zero and in the severely ill–posed case of positive volatility. We treat both cases. A confidence set is called honest if the level is achieved uniformly over a class of the estimated objects. We prove asymptotic normality uniformly over a class of characteristic triplets and use this to construct honest confidence intervals. The asymptotic normality results are based on undersmoothing and on a linearization of the stochastic errors. As it turns out the asymptotic distribution is completely determined by the linearized stochastic errors.

Based on the asymptotic analysis, we construct confidence intervals from the finite sample variance of the linearized stochastic errors. We study the performance of the confidence intervals in simulations and apply the confidence intervals to option data of the German DAX index. While we focus in this thesis on the spectral calibration method by Belomestny and Reiß (2006a), the approach can be easily generalized to similar methods. The construction of the confidence sets, the simulations and the empirical study will appear in Söhl and Trabs (2012b), where this is also carried out for the method by Trabs (2012).

Nonparametric confidence intervals and sets for jump densities have been studied by Figueroa-López (2011). The method is based on direct high–frequency observations so that the statistical problem of estimating the jump density is easier than in our set–up. On the other hand the results yield beyond pointwise confidence intervals also confidence bands.

For low–frequency observations, Nickl and Reiß (2012) show in a recent paper a central limit theorem for the nonparametric estimation of Lévy processes. They consider the estimation of the generalized distribution function of the Lévy measure in the mildly ill–posed case for particular situations in which a $\sqrt{n}$–rate can be obtained and prove a uniform central limit theorem for their estimators.

So while central limit theorems have been treated in the nonparametric estimation of Lévy processes for low–frequency observations in the mildly ill–posed case, there are to the best of the author's knowledge no results in the severely ill–posed case and our results are the first for this case.

The estimation of the characteristic triplet from low–frequency observations of a Lévy processes is closely related to the deconvolution problem. Considering the deconvolution problem for two different densities f_X and $\overline{f_X}$ yields equation (8.2), namely

$$\mathcal{F}[\overline{f_X} - f_X](u) = \frac{\overline{\varphi}(u) - \varphi(u)}{\varphi_\varepsilon(u)}, \tag{1.1}$$

where φ and $\overline{\varphi}$ are the characteristic functions of the observations belonging to f_X and $\overline{f_X}$, respectively, and φ_ε is the characteristic function of the errors ε_j. The corresponding formula (2.28) for two characteristic triplets of a Lévy process exhibits the same structure. The difference is that there φ and φ_ε are both replaced by the same characteristic function φ_T of the Lévy process. This is called auto–deconvolution, since the distribution of the errors is replaced by the marginal distribution of the Lévy process itself. From the structure of the above formula one can also see that the decay of the characteristic functions φ_ε and φ_T, respectively, determines the ill–posedness of the problems. A polynomial decay corresponds to the mildly ill–posed case and an exponential decay to the severely ill–posed case.

We consider the deconvolution problem in the mildly ill–posed case, which is the deconvolution problem with ordinary smooth errors. Using kernel estimators, we estimate linear functionals $\int \zeta(x-t) f_X(x)\,\mathrm{d}x$, which are general enough to include the estimation of the distribution function as a special case. Our main result in the deconvolution problem is a central limit theorem for the estimators uniformly over all $t \in \mathbb{R}$. This result will appear in Söhl and Trabs (2012a), where in addition also the efficiency of the estimators is shown. Similarly to the situation considered by Nickl and Reiß (2012) for Lévy processes, we treat the case where a $\sqrt{n}$–rate can be obtained. Our work gives a clear insight into the interplay between the smoothness of ζ and the ill–posedness of the problem. A $\sqrt{n}$–rate can be obtained whenever the smoothness of ζ in an L^2–Sobolev sense compensates the ill–posedness of the problem determined by the polynomial rate by which the characteristic function of the error decays. The limit process $\mathbb{G}$ in the uniform central limit theorem is a generalized Brownian bridge, whose covariance depends on the functional ζ and through the deconvolution operator $\mathcal{F}^{-1}[1/\varphi_\varepsilon]$ also on the distribution of the error. By uniform convergence the kernel estimator of f_X fulfills the 'plug–in' property of Bickel and Ritov (2003). The theory of smoothed empirical processes as treated in Radulović and Wegkamp (2000) as well as in Giné and Nickl (2008) is used to prove the uniform central limit theorem.

Deconvolution is a well studied problem. So we focus here only on the closely related literature and refer to the references therein for further reading. Fan (1991b) treats minimax convergence rates for estimating the density and the distribution function. Butucea and Comte (2009) treat the data–driven choice of the bandwidth for estimating linear functionals of the density f_X, but assume some minimal smoothness and integrability conditions on the functional, which exclude, for example, the estimation of the distribution function since $\mathbf{1}_{(-\infty,t]}$ is not integrable. Dattner et al. (2011) study the minimax–optimal and adaptive estimation of the distribution function.

In view of our work, we focus now on asymptotic normality and on confidence sets. Deconvolution is generally considered in the mildly ill–posed case of ordinary smooth errors and in the severely ill–posed case of supersmooth errors. In ordinary smooth deconvolution, asymptotic normality is shown for the estimation of the density by Fan (1991a) and on slightly weaker assumptions by Fan and Liu (1997). For the estimation of the distribution function, Hall and Lahiri (2008) show asymptotic normality in ordinary smooth deconvolution. In supersmooth deconvolution, asymptotic normality is proved for estimators of the density by Zhang (1990) and by Fan (1991a). Zhang (1990) covers also estimators of the distribution function and van Es and Uh (2005) further determine the asymptotic behavior of the variance for estimators of the density and of the distribution function in supersmooth deconvolution. The asymptotic normality results on supersmooth deconvolution are extended by van Es and Uh (2004) to the case when the characteristic function of the errors decays exponentially but possibly slower than the one of the Cauchy distribution. Further developing the work by Bickel and Rosenblatt (1973) on density estimation, Bissantz et al. (2007) construct confidence bands for the density in ordinary smooth deconvolution. Lounici and Nickl (2011) give uniform risk bounds for wavelet density estimators in the deconvolution problem, which can be used to construct nonasymptotic confidence bands.

In both problems, we apply spectral regularization. The higher the frequencies, the more they contribute to the stochastic error. We regularize by discarding all frequencies higher than a certain cut–off value. Since we regularize in the spectral domain, Fourier techniques are used for the estimation methods and their analysis.

Another common feature is that Gaussian processes arise naturally in both problems. The limit process of the stochastic error in the deconvolution problem is a generalized Brownian bridge. The problem of estimating the characteristic triplet of a Lévy process can be simplified by studying observations in the Gaussian white noise model. Applying the estimation method to this modified observation scheme leads to a Gaussian process. A bound on the supremum of this Gaussian process is derived which is later used to prove asymptotic normality of the estimators. For the Gaussian processes in both problems, we study boundedness and continuity using Dudley's theorem and metric entropy arguments. While these are classical topics in the theory of Gaussian processes, we also address the question of hitting probabilities for Gaussian processes or, more generally, for Gaussian random fields. These results on hitting probabilities are of independent interest and are published in Söhl (2010). They are used to show that points are polar for the Gaussian process resulting from the Gaussian white noise model meaning that it does not hit a given point almost surely. This implies that the estimators in the Lévy setting are almost surely well–defined.

In both problems, in the deconvolution and in the estimation of the Lévy process, we use nonparametric estimation methods. The estimation errors can be decomposed into a stochastic and an approximation part. Unlike the bias–variance trade–off suggests, we do not try to balance stochastic and approximation error but rather aim for undersmoothing. Then the approximation error is asymptotically negligible and thus the asymptotic distribution is centered around the true value. The asymptotic variance can be easily estimated by means of the already used estimators. In contrast to a bias correction, which often leads to more difficult estimation problems, undersmoothing yields accessible asymptotic distributions and feasible confidence sets.

This thesis is organized as follows. Chapter 2 treats the exponential Lévy model and the spectral estimation method. Chapter 3 studies continuity, boundedness and hitting probabilities of a related Gaussian process. Chapter 4 contains the main results on asymptotic normality in the Lévy setting. Chapter 5 treats uniform convergence with respect to the underlying probability measure. In Chapter 6 the asymptotic normality results are applied to confidence sets and to a hypotheses test on the value of the volatility. Chapter 7 contains a finite sample analysis, simulations and an empirical study on the calibration of the exponential Lévy model. Chapter 8 treats the deconvolution model and is devoted to a uniform central limit theorem for estimators of linear functionals of the density. We conclude and give an outlook on further research topics in Chapter 9.

2 Calibration of exponential Lévy models

This chapter introduces the spectral calibration method and begins to analyze the estimation error. To this end, we provide some background on Lévy processes in Section 2.1. We describe the spectral calibration method and a slight modification thereof both by Belomestny and Reiß (2006a,b) in Section 2.2, where we also briefly discuss the structural similarity of the calibration and the deconvolution problem. In Section 2.3, an example of model misspecification is considered. We introduce an error decomposition in Section 2.4 which will be important later for the further analysis of the errors.

2.1 Lévy processes

In this section, we define Lévy processes and summarize some of their properties, which can be found, for example, in the monograph by Sato (1999). Later we will need only one dimensional Lévy processes. Nevertheless, we treat here Lévy processes with values in $\mathbb{R}^d$ since this causes no additional effort.

Definition 2.1 (Lévy process)**.** An $\mathbb{R}^d$–valued stochastic process $(L_t)_{t\geqslant 0}$ on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *Lévy process* if the following properties are satisfied:

(i) $L_0 = 0$ almost surely,

(ii) (L_t) has independent increments: for any choice of $n \geqslant 1$ and $0 \leqslant t_0 < t_1 < \cdots < t_n$ the random variables $L_{t_0}, L_{t_1} - L_{t_0}, \ldots, L_{t_n} - L_{t_{n-1}}$ are independent,

(iii) (L_t) has stationary increments: the distribution of $L_{t+s} - L_t$ does not depend on t,

(iv) (L_t) is stochastically continuous: for all $t \geqslant 0$, $\varepsilon > 0$, $\lim_{s\to 0} \mathbb{P}(|L_{t+s} - L_t| > \varepsilon) = 0$,

(v) (L_t) has almost surely càdlàg paths: there exists $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that for all $\omega \in \Omega_0$, $L_t(\omega)$ is right–continuous at all $t \geqslant 0$ and has left limits at all $t > 0$.

Example 2.2. (i) A Brownian motion with a deterministic drift $(\Sigma B_t + \gamma t)$ is a Lévy process, where $\Sigma \in \mathbb{R}^{d\times d}$, $\gamma \in \mathbb{R}^d$ and B_t is a d–dimensional Brownian motion.

(ii) A Poisson process (N_t) of intensity $\lambda \geqslant 0$ is a Lévy process. More generally, the compound Poisson process (Y_t) is a Lévy process, where $Y_t := \sum_{j=1}^{N_t} Z_j$ with independent, identically distributed random variables Z_j, which take values in $\mathbb{R}^d$.

We note that the sum of two independent Lévy processes is again a Lévy process. These examples capture the behavior of Lévy processes quite well. Indeed, the Lévy–Itô decomposition states that any Lévy process can be represented as the sum of three independent components $L = L_1 + L_2 + L_3$, where L_1 is a Brownian motion with deterministic drift, L_2 is a compound Poisson process with jumps larger or equal to one and L_3 is a martingale representing the possible infinitely many jumps smaller than one, which may be obtained as the limit of compensated compound Poisson processes with jumps smaller than one. We call L_1 the *continuous part* and $L_2 + L_3$ the *jump part* of the Lévy process. For a precise formulation of the Lévy–Itô decomposition and a proof we refer to Sato (1999). Another main result on Lévy processes is the Lévy–Khintchine representation, whose statement and proof can be found in the same monograph.

Theorem 2.3 (Lévy–Khintchine representation)**.** *Let (L_t) be a Lévy process. Then there exists a unique triplet (A, b, ν) consisting of a symmetric positive semi–definite matrix $A \in \mathbb{R}^{d\times d}$, a vector $b \in \mathbb{R}^d$ and a measure ν on $\mathbb{R}^d$ satisfying*

$$\nu(\{0\}) = 0 \quad \textit{and} \quad \int_{\mathbb{R}^d} (|x|^2 \wedge 1)\nu(\,\mathrm{d}x) < \infty, \tag{2.1}$$

such that for all $T \geqslant 0$ and for all $u \in \mathbb{R}^d$

$$\begin{aligned}\varphi_T(u) &:= \mathbb{E}[e^{i\langle u, L_T\rangle}] \\ &= \exp\left(T\left(-\frac{1}{2}\langle u, Au\rangle + i\langle b, u\rangle + \int_{\mathbb{R}^d} (e^{i\langle u,x\rangle} - 1 - i\langle u, x\rangle \mathbf{1}_{\{|x|\leqslant 1\}}(x))\nu(\,\mathrm{d}x)\right)\right)\end{aligned}$$

Conversely, let (A, b, ν) be a triplet consisting of a symmetric positive semi–definite matrix $A \in \mathbb{R}^{d\times d}$, a vector $b \in \mathbb{R}^d$ and a measure ν on $\mathbb{R}^d$ satisfying the properties (2.1), *then there exists a Lévy process with characteristic function given by the above equation.*

We call $A \in \mathbb{R}^{d\times d}$ the *Gaussian covariance matrix* and ν the *Lévy measure*. The *Blumenthal–Getoor index* is defined as

$$\alpha := \inf\left\{ r \geqslant 0 \,\middle|\, \int_{|x|\leqslant 1} |x|^r \nu(\,\mathrm{d}x) < \infty \right\}$$

and measures the degree of activity of the small jumps. The jump part of a Lévy process is almost surely of bounded variation on compact sets if and only if $\int_{\{|x|\leqslant 1\}} |x|\nu(\,\mathrm{d}x) < \infty$. In this case we define the *drift* $\gamma := b - \int_{\{|x|\leqslant 1\}} x\nu(\,\mathrm{d}x) \in \mathbb{R}$. (A, γ, ν) is called the *characteristic triplet* of the Lévy process (L_t). If the *intensity* $\lambda := \nu(\mathbb{R}^d)$ is finite, then the jump part is a compound Poisson process. For a one dimensional Lévy process we write σ^2 instead of A and call $\sigma \geqslant 0$ the *volatility.*

2.2 Spectral calibration method

In this section, we introduce the exponential Lévy model and describe the spectral calibration method. A slight modification of the spectral calibration method is also explained. At the end of this section, we discuss the similarity of the calibration and the deconvolution problem.

Exponential Lévy models describe the price of an asset by

$$S_t = Se^{rt+L_t} \quad \text{with a } \mathbb{R}\text{–valued Lévy process } (L_t) \text{ for } t \geqslant 0. \tag{2.2}$$

A thorough discussion of this model is given in the monograph by Cont and Tankov (2004a). Since the method is based on option prices, the calibration is in the risk neutral world modeled by a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t))$. We assume that the discounted price process is a martingale with respect to the risk–neutral measure $\mathbb{P}$ and that under $\mathbb{P}$ the price of the asset (S_t) follows the exponential Lévy model (2.2), where $S > 0$ is the present value of the asset and $r \geqslant 0$ is the riskless interest rate.

A European call option with strike price K and maturity T is the right but not the obligation to buy an asset for price K at time T. A European put option is the respective right for selling the asset. We denote by $\mathcal{C}(K,T)$ and $\mathcal{P}(K,T)$ the prices of European call and put options which are determined by the pricing formulas

$$\mathcal{C}(K,T) = e^{-rT}\,\mathbb{E}[(S_T - K)^+], \tag{2.3}$$

$$\mathcal{P}(K,T) = e^{-rT}\,\mathbb{E}[(K - S_T)^+], \tag{2.4}$$

where we used the notion $(A)^+ := \max(A, 0)$. Subtracting (2.4) from (2.3) yields the well known put–call parity

$$\mathcal{C}(K,T) - \mathcal{P}(K,T) = S - e^{-rT}K,$$

where we used that $(e^{-rt}S_t)$ is a martingale. By the put–call parity, call prices can be calculated into put prices and vice versa so that the observation may be given by either of them. We fix some T and suppose that the observed option prices correspond to different maturities (K_j) and are given by the value of the pricing formula corrupted by noise:

$$Y_j = \mathcal{C}(K_j, T) + \eta_j \xi_j, \quad j = 1, \ldots, n. \tag{2.5}$$

The minimax result in Belomestny and Reiß (2006a) is shown for general errors (ξ_j) which are independent, centered random variables with $\mathrm{Var}(\xi_j) = 1$ and $\sup_j \mathbb{E}[\xi_j^4] < \infty$. The observation errors are due to the bid–ask spread and other market frictions. The noise levels (η_j) can be either determined from the bid–ask spread, which indicates by Cont and Tankov (2004a, p. 438/439) how reliable an observation is, or they can be estimated nonparametrically, for example, with the method by Fan and Yao (1998). We transform the observations to a regression problem on the function

$$\mathcal{O}(x) := \begin{cases} S^{-1}\mathcal{C}(x,T), & x \geqslant 0, \\ S^{-1}\mathcal{P}(x,T), & x < 0, \end{cases}$$

where $x := \log(K/S) - rT$ denotes the negative log–forward moneyness. The regression model may then be written as

$$O_j = \mathcal{O}(x_j) + \delta_j \xi_j, \tag{2.6}$$

where $\delta_j = S^{-1}\eta_j$. Since the design may change with n, it would be more precise to index the regression model (2.6) by nj instead of by j only. But for notational convenience we omit the dependence on n.

Denoting by δ_0 the dirac measure at zero, we define the measure $\nu_\sigma(\mathrm{d}x) := \sigma^2\delta_0(\mathrm{d}x) + x^2\nu(\mathrm{d}x)$. Its structure in a neighborhood of zero is very natural, since it is most useful in characterizing weak convergence of the distribution of the Lévy process in view of Theorem VII.2.9 and Remark VII.2.10 in Jacod and Shiryaev (2003). The measure ν_σ determines the variance of a Lévy process and is relevant for calculating the Δ in quadratic hedging as noted in Neumann and Reiß (2009). Volatility and small jumps both contribute to the mass assigned by ν_σ to a neighborhood of zero. Thus it is difficult to distinguish between small jumps and volatility, in fact Neumann and Reiß (2009) point out in their Remark 3.2 that without further restrictions the volatility cannot be estimated consistently. As in Belomestny and Reiß (2006a) we will resolve this problem by considering only Lévy processes (L_t) with finite intensity and with a Lévy measure, which is absolutely continuous with respect to the Lebesgue measure. Since then the Lévy measure is determined by its Lebesgue density, we will denote the Lebesgue density in the following likewise by ν. Then the Lévy–Khintchine representation in Theorem 2.3 simplifies to

$$\varphi_T(u) := \mathbb{E}[e^{iuL_T}] = \exp\left(T\left(-\frac{\sigma^2u^2}{2} + i\gamma u + \int_{-\infty}^{\infty}(e^{iux}-1)\nu(x)\,\mathrm{d}x\right)\right), \tag{2.7}$$

where we call $\sigma \geqslant 0$ the *volatility*, $\gamma \in \mathbb{R}$ the *drift* and $\nu \in L^1(\mathbb{R})$ the *jump density* with *intensity* $\lambda := \|\nu\|_{L^1(\mathbb{R})}$. We call (σ^2, γ, ν) the *characteristic triplet* of the Lévy process (L_t).

In view of the Lévy–Khintchine representation (2.7) and of the independent increments, the martingale property of $(e^{-rt}S_t)$ may be equivalently characterized by

$$\mathbb{E}[e^{L_T}] = 1\ \forall T \geqslant 0 \quad \Longleftrightarrow \quad \frac{\sigma^2}{2} + \gamma + \int_{-\infty}^{\infty}(e^x-1)\nu(x)\,\mathrm{d}x = 0. \tag{2.8}$$

For a jump density ν we denote by $\mu(x) := e^x\nu(x)$ the corresponding exponentially weighted jump density. The aim is to estimate the characteristic triplet $\mathcal{T} = (\sigma^2, \gamma, \mu)$ (we use both equivalent parametrization of characteristic triplets in μ and ν). We will assume that $\mathbb{E}[e^{2L_T}]$ is finite, which implies that there is a constant $C > 0$ such that $\mathcal{O}(x) \leqslant Ce^{-|x|}$ for all $x \in \mathbb{R}$ (Belomestny and Reiß, 2006a, Prop. 2.1). This assumption is equivalent to assuming a finite second moment of the asset price, $\mathbb{E}[S_T^2] < \infty$. Then $\mathcal{O}$ is integrable and we can consider the Fourier transform $\mathcal{F}\mathcal{O}$.

In the remainder of this section we present and discuss the spectral calibration method and a slight modification thereof both by Belomestny and Reiß (2006a,b). The method is based on the Lévy–Khintchine representation (2.7) and on an option pricing formula by Carr and Madan (1999)

$$\mathcal{F}\mathcal{O}(u) := \int_{-\infty}^{\infty} e^{iux}\mathcal{O}(x)\,\mathrm{d}x = \frac{1-\varphi_T(u-i)}{u(u-i)}, \tag{2.9}$$

which holds on the strip $\{u \in \mathbb{C} \,|\, \mathrm{Im}(u) \in [0,1]\}$. We define

$$\begin{aligned}\psi(u) &:= \frac{1}{T}\log\left(1+iu(1+iu)\mathcal{F}\mathcal{O}(u)\right) = \frac{1}{T}\log(\varphi_T(u-i))\\ &= -\frac{\sigma^2u^2}{2} + i(\sigma^2+\gamma)u + (\sigma^2/2+\gamma-\lambda) + \mathcal{F}\mu(u),\end{aligned} \tag{2.10}$$

where the first equality is given by the pricing formula (2.9) and the second by the Lévy–Khintchine representation (2.7). The second equality holds likewise on the strip $\{u \in \mathbb{C} \,|\, \mathrm{Im}(u) \in [0,1]\}$ since there the characteristic function $\varphi_T(u-i)$ is finite by the exponential moment of L_T in the martingale condition (2.8). This equation links the observations of $\mathcal{O}$ to the characteristic triplet that we want to estimate. Let $\mathcal{O}_n$ be an empirical version of the true function $\mathcal{O}$. For example, $\mathcal{O}_n$ can be obtained by linear interpolation of the data (2.6). Replacing $\mathcal{O}$ by $\mathcal{O}_n$ in equation (2.10) yields an empirical version of ψ. For direct observations of the Lévy process at low frequency one could plug in the empirical characteristic function of the observations into equation (2.10) and then proceed as described below. But we will stick to the observations in terms of option prices. So we define the empirical counterpart of ψ by

$$\psi_n(u) := \frac{1}{T}\log_{\geqslant\kappa(u)}\left(1+iu(1+iu)\mathcal{F}\mathcal{O}_n(u)\right), \tag{2.11}$$

where the trimmed logarithm $\log_{\geqslant\kappa} : \mathbb{C}\setminus\{0\} \to \mathbb{C}$ is given by

$$\log_{\geqslant\kappa}(z) := \begin{cases} \log(z), & |z| \geqslant \kappa\\ \log(\kappa\, z/|z|), & |z| < \kappa\end{cases}$$

and $\kappa(u) := \exp(-T\sigma_{\max}^2u^2/2 - 4TR)/2 \in (0,1)$. The quantities $\sigma_{\max}, R \geqslant 0$ are determined by the class of characteristic triplets in Definition 2.4 below. The logarithms are taken in such a way that ψ and ψ_n are continuous with $\psi(0) = \psi_n(0) = 0$. This way of taking the complex logarithm is called *distinguished logarithm*, see Theorem 7.6.2 in Chung (1974). We will further discuss the distinguished logarithm especially in connection with the definition of ψ_n in Chapter 3. Considering (2.10) as a quadratic polynomial in u disturbed by $\mathcal{F}\mu$ motivates the following definitions of the estimators for a cut–off value $U > 0$:

$$\hat{\sigma}^2 := \int_{-U}^{U}\mathrm{Re}(\psi_n(u))w_\sigma^U(u)\mathrm{d}u, \tag{2.12}$$

$$\hat{\gamma} := -\hat{\sigma}^2 + \int_{-U}^{U}\mathrm{Im}(\psi_n(u))w_\gamma^U(u)\mathrm{d}u, \tag{2.13}$$

$$\hat{\lambda} := \frac{\hat{\sigma}^2}{2} + \hat{\gamma} - \int_{-U}^{U} \mathrm{Re}(\psi_n(u)) w_\lambda^U(u) \mathrm{d}u, \tag{2.14}$$

where the weight functions w_σ^U, w_γ^U and w_λ^U satisfy

$$\begin{aligned} &\int_{-U}^{U} \frac{-u^2}{2} w_\sigma^U(u) \mathrm{d}u = 1, \quad \int_{-U}^{U} u w_\gamma^U(u) \mathrm{d}u = 1, \quad \int_{-U}^{U} w_\lambda^U(u) \mathrm{d}u = 1, \\ &\int_{-U}^{U} w_\sigma^U(u) \mathrm{d}u = 0, \quad \int_{-U}^{U} u^2 w_\lambda^U(u) \mathrm{d}u = 0. \end{aligned} \tag{2.15}$$

The estimator for μ is defined by a smoothed inverse Fourier transform of the remainder

$$\hat{\mu}(x) := \mathcal{F}^{-1}\left[\left(\psi_n(u) + \frac{\hat{\sigma}^2}{2}(u-i)^2 - i\hat{\gamma}(u-i) + \hat{\lambda}\right) w_\mu^U(u)\right](x), \tag{2.16}$$

where w_μ^U is compactly supported. The choice of the weight functions is discussed in Section 7.1, where also possible weight functions are given. The weight functions for all $U > 0$ can be obtained from w_σ^1, w_γ^1, w_λ^1 and w_μ^1 by rescaling:

$$\begin{aligned} w_\sigma^U(u) &= U^{-3} w_\sigma^1(u/U), \quad w_\gamma^U(u) = U^{-2} w_\gamma^1(u/U), \\ w_\lambda^U(u) &= U^{-1} w_\lambda^1(u/U), \quad w_\mu^U(u) = w_\mu^1(u/U). \end{aligned}$$

Since $\psi_n(-u) = \overline{\psi_n(u)}$, only the symmetric part of w_σ^1, w_λ^1 and the antisymmetric part of w_γ^1 matter. The antisymmetric part of w_μ^1 contributes a purely imaginary part to $\hat{\mu}(x)$. Without loss of generality we will always assume w_σ^1, w_λ^1, w_μ^1 to be symmetric and w_γ^1 to be antisymmetric. We further assume that the supports of w_σ^1, w_γ^1, w_λ^1 and w_μ^1 are contained in $[-1, 1]$.

To bound the approximation errors some smoothness assumption is necessary. We assume that the characteristic triplet belongs to a smoothness class given by the following definition, which is Definition 4.1 by Belomestny and Reiß (2006a).

Definition 2.4. For $s \in \mathbb{N}$ and $R, \sigma_{\max} \geqslant 0$ let $\mathcal{G}_s(R, \sigma_{\max})$ denote the set of all characteristic triplets $\mathcal{T} = (\sigma^2, \gamma, \mu)$ such that (e^{L_t}) is a martingale, $\mathbb{E}[e^{2L_T}] \leqslant R$ holds, μ is s–times (weakly) differentiable and

$$\sigma \in [0, \sigma_{\max}], \quad |\gamma|, \lambda \in [0, R], \quad \max_{0 \leqslant k \leqslant s} \|\mu^{(k)}\|_{L^2(\mathbb{R})} \leqslant R, \quad \|\mu^{(s)}\|_{L^\infty(\mathbb{R})} \leqslant R.$$

The assumption $\mathcal{T} \in \mathcal{G}_s(R, \sigma_{\max})$ includes a smoothness assumption of order s on μ leading to a decay of $\mathcal{F}\mu$. To profit from this decay when bounding the approximation error in Section 4.3.3, we assume that the weight functions are of order s, this means

$$\mathcal{F}(w_\sigma^1(u)/u^s), \mathcal{F}(w_\gamma^1(u)/u^s), \mathcal{F}(w_\gamma^1(u)/u^s), \mathcal{F}((1 - w_\mu^1(u))/u^s) \in L^1(\mathbb{R}). \tag{2.17}$$

A slightly modified estimation method is given in a second paper by Belomestny and Reiß (2006b), which is concerned with simulations and an empirical example. We present this approach here for completeness and since our simulations and our empirical study are partly based on it. In addition, we will apply the modified method to an example and this will shed some light on misspecification in the next section. As noted, the equation (2.10) holds on the whole strip $\{u \in \mathbb{C} \,|\, \mathrm{Im}(u) \in [0,1]\}$. So instead of using this equation directly one could also shift the equation. A shift by one in the imaginary direction is particularly appealing, since then ν can be estimated directly without the intermediate step of estimating an exponentially scaled version of ν. Applying this shift to (2.10) yields

$$\begin{aligned}\psi(u+i) &= \frac{1}{T}\log(1-u(u+i)\,\mathcal{F}\,\mathcal{O}(u+i)) = \frac{1}{T}\log(\varphi_T(u)) \\ &= -\frac{\sigma^2 u^2}{2} + i\gamma u - \lambda + \mathcal{F}\,\nu(u).\end{aligned} \tag{2.18}$$

Similar as for equation (2.10), one could also plug in an empirical characteristic function obtained from direct, low–frequency observations here. This is exactly the approach Gugushvili (2009) takes. But we notice $\mathcal{F}\,\mathcal{O}(u+i) = \mathcal{F}[e^{-x}\mathcal{O}(x)](u)$. Again we substitute $\mathcal{O}$ by its empirical counterpart $\mathcal{O}_n$ and define

$$\psi_n(u+i) := \frac{1}{T}\log(1-u(u+i)\,\mathcal{F}[e^{-x}\mathcal{O}_n(x)](u)). \tag{2.19}$$

So the slightly modified estimators are given by

$$\widetilde{\sigma}^2 := \int_{-U}^{U} \mathrm{Re}(\psi_n(u+i))w_\sigma^U(u)\mathrm{d}u, \tag{2.20}$$

$$\widetilde{\gamma} := \int_{-U}^{U} \mathrm{Im}(\psi_n(u+i))w_\gamma^U(u)\mathrm{d}u, \tag{2.21}$$

$$\widetilde{\lambda} := -\int_{-U}^{U} \mathrm{Re}(\psi_n(u+i))w_\lambda^U(u)\mathrm{d}u, \tag{2.22}$$

where the weight functions are assumed to satisfy the same conditions (2.15) as before. The corresponding estimator of ν is

$$\widetilde{\nu}(x) := \mathcal{F}^{-1}\left[\left(\psi_n(u+i) + \frac{\widetilde{\sigma}^2}{2}u^2 - i\widetilde{\gamma}u + \widetilde{\lambda}\right)w_\nu^U(u)\right](x), \tag{2.23}$$

where w_ν^U is compactly supported. Belomestny and Reiß (2006b) proposed the weight functions

$$w_\sigma^U(u) := \frac{s+3}{1-2^{-2/(s+1)}}U^{-(s+3)}|u|^s(1-2\cdot\mathbf{1}_{\{|u|>2^{-1/(s+1)}U\}}), \quad u\in[-U,U], \tag{2.24}$$

$$w_\gamma^U(u) := \frac{s+2}{2U^{s+2}}|u|^s\,\mathrm{sgn}(u), \quad u\in[-U,U], \tag{2.25}$$

$$w_\lambda^U(u) := \frac{s+1}{2(2^{2/(s+3)}-1)}U^{-(s+1)}|u|^s(2\cdot\mathbf{1}_{\{|u|<2^{-1/(s+3)}U\}}-1), \quad u\in[-U,U], \tag{2.26}$$

$$w_\nu^U(u) := (1-(u/U)^2)^+, \quad u \in \mathbb{R}\,. \tag{2.27}$$

In order to see the similarity of the estimation of the characteristic triplet of a Lévy process with the deconvolution problem we rewrite (2.18) as

$$T\,\mathcal{F}((\sigma^2/2)\delta_0'' - \gamma\delta_0' - \lambda\delta_0 + \nu)(u) = \log(\varphi_T(u)).$$

In the spectral calibration method φ_T is replaced by an empirical version and then this equation is used to estimate the characteristic triplet. We consider the equation for two different characteristic triplets (σ,γ,ν) and $(\overline{\sigma},\overline{\gamma},\overline{\nu})$ with intensities λ and $\overline{\lambda}$, respectively, and obtain

$$\begin{aligned} &T\,\mathcal{F}(((\overline{\sigma}^2-\sigma^2)/2)\delta_0'' - (\overline{\gamma}-\gamma)\delta_0' - (\overline{\lambda}-\lambda)\delta_0 + (\overline{\nu}-\nu))(u) \\ &\quad = \log(\overline{\varphi}_T(u)) - \log(\varphi_T(u)) \\ &\quad = \log\left(1 + \frac{\overline{\varphi}_T(u)-\varphi_T(u)}{\varphi_T(u)}\right) \approx \frac{\overline{\varphi}_T(u)-\varphi_T(u)}{\varphi_T(u)}, \end{aligned} \tag{2.28}$$

where the approximation is valid if the absolute value of the last expression is small. This formula reveals the deconvolution structure of the problem. On the one hand the similarity with the corresponding formula for deconvolution (1.1) is striking. On the other hand we can see the deconvolution structure directly from formula (2.28). To this end, we multiply with φ_T on both sides. Since a multiplication in the spectral domain corresponds to a convolution in the spatial domain, we see that the difference in the characteristic triplet is convolved with the marginal distribution of the Lévy process. To estimate the characteristic triplet a deconvolution problem has to be solved. Interestingly, the marginal distribution of the Lévy process appears twice, it takes the place of both, the error distribution and the distribution of the observations in the deconvolution problem. This phenomenon is called *auto–deconvolution*.

The linearization of the logarithm in (2.28) will be important later, when we substitute the logarithm (2.31) in the stochastic errors by its linearization (2.32). Then $\overline{\varphi}_T$ will be an empirical version of the characteristic function φ_T, which justifies the assumption that $\overline{\varphi}_T$ and φ_T are close. The division by the possibly decaying function φ_T is taken care of in the estimation method by the spectral cut–off.

2.3 The misspecified model

We have chosen a nonparametric estimation method to reduce the error due to model misspecification and assume in general that the misspecification error is negligible. Nevertheless, model misspecification is an important issue to address. In this section, we want to study at least by means of an example how the misspecified model behaves. The spectral calibration method is designed for finite intensity Lévy processes. Sudden changes in the price process are incorporated into the model by jumps of the Lévy process. The Gaussian component models the small fluctuations happening all the time. Alternatively, one can interpret these fluctuations to be caused by infinitely many small jumps. This can be modeled by an infinite intensity Lévy process and empirical investigations indicate that Lévy processes with Blumenthal–Getoor index larger than one are particularly suitable. Stable processes allow to consider different Blumenthal–Getoor indices $\alpha \in (0,2)$. So we consider a symmetric stable process with additional drift and Gaussian component and study the behavior of the estimators for such a process. The characteristic function is given by

$$\varphi_T(u) = \exp(T(-\sigma^2 u^2/2 - \eta^\alpha |u|^\alpha + i\gamma u)),$$

where $\alpha \in (0,2)$, $\sigma, \eta \geqslant 0$ and $\gamma \in \mathbb{R}$. It holds $\psi(u+i) = -\sigma^2 u^2/2 - \eta^\alpha |u|^\alpha + i\gamma u$. We take the weight functions w_σ^U, w_γ^U and w_λ^U as in (2.24), (2.25) and (2.26), respectively, and $w_\nu^U = \mathbf{1}_{[-U,U]}$. We apply the second method given by (2.20)–(2.23) directly to $\psi(u+i)$ and not to its empirical counterpart $\psi_n(u+i)$ and obtain

$$\begin{aligned}
\widetilde{\sigma}^2 &= \int_{-U}^{U} \operatorname{Re}(\psi(u+i)) w_\sigma^U(u)\mathrm{d}u = \sigma^2 + \frac{2 - 2^{(s+1-\alpha)/(s+1)}}{1 - 2^{-2/(s+1)}} \frac{s+3}{s+\alpha+1} \eta^\alpha U^{-(2-\alpha)}, \\
\widetilde{\gamma} &= \int_{-U}^{U} \operatorname{Im}(\psi(u+i)) w_\gamma^U(u)\mathrm{d}u = \gamma, \\
\widetilde{\lambda} &= -\int_{-U}^{U} \operatorname{Re}(\psi(u+i)) w_\lambda^U(u)\mathrm{d}u = \frac{2^{(2-\alpha)/(s+3)} - 1}{2^{2/(s+3)} - 1} \frac{s+1}{s+\alpha+1} \eta^\alpha U^\alpha, \\
\widetilde{\nu}(0) &= \mathcal{F}^{-1}[(\psi(u+i) + \widetilde{\sigma}^2 u^2/2 - i\widetilde{\gamma} u + \widetilde{\lambda})\mathbf{1}_{[-U,U]}(u)](0) \\
&= (-\eta^\alpha U^{\alpha+1}/(\alpha+1) + (\widetilde{\sigma}^2 - \sigma^2)U^3/6 + \widetilde{\lambda} U)/\pi.
\end{aligned}$$

We observe that the drift γ is estimated correctly. The estimated volatility $\widetilde{\sigma}^2$ converges with rate $U^{-(2-\alpha)}$ to σ^2. The estimated jump intensity is finite, but grows as U^α. The estimated jump intensity at zero $\widetilde{\nu}(0)$ grows as $U^{\alpha+1}$. Although the estimated jump intensities are always finite, the infinite intensity of the Lévy process is reflected in the estimators by growing jump intensities and by peaks of growing height at zero. For a Lévy process of infinite jump intensity, $\widetilde{\sigma}^2$ has to be interpreted as a joint quantity of volatility and small jumps. The corresponding singularity of the jump density is given by $c\eta^\alpha/|x|^{\alpha+1}$ with $c > 0$. The smoothing by the weight function in the spectral domain corresponds to a kernel smoothing with bandwidth U^{-1} in the spatial domain. The integral of $x^2\nu(x)$ in a neighborhood of zero with size U^{-1} is proportional to

$$\int_{-U^{-1}}^{U^{-1}} x^2 \nu(x)\mathrm{d}x = \int_{-U^{-1}}^{U^{-1}} c\eta^\alpha |x|^{1-\alpha}\mathrm{d}x = \frac{2}{2-\alpha} c\eta^\alpha U^{\alpha-2}.$$

We see that $\widetilde{\sigma}^2$ behaves as the mass assigned by ν_σ to a neighborhood of zero with size proportional to U^{-1}. This example shows that with the above interpretation the estimators can give valuable information about the process even in the case of model misspecification.

2.4 Preliminary error analysis

In this section, we start with a preliminary analysis of the error in the correctly specified model. We decompose the error into an approximation error and a stochastic error. The stochastic error is further decomposed into a linearized part and a remainder term. The linearized stochastic error is considered in the Gaussian white noise model. This leads to the definition of the Gaussian process studied in Chapter 3. At the same time this error decomposition lays the ground for the proof of the asymptotic normality in Chapter 4.

We define the estimation error $\Delta\hat{\sigma}^2 := \hat{\sigma}^2 - \sigma^2$ and likewise for the other estimators. We will also use the notation $\Delta\psi_n := \psi_n - \psi$. The estimation error $\Delta\hat{\sigma}^2$ can be decomposed as

$$\Delta\hat{\sigma}^2 := \frac{2}{U^2}\int_0^1 \mathrm{Re}(\mathcal{F}\mu(Uu))w_\sigma^1(u)\mathrm{d}u + \frac{2}{U^2}\int_0^1 \mathrm{Re}(\Delta\psi_n(Uu))w_\sigma^1(u)\mathrm{d}u. \tag{2.29}$$

The first term is the approximation error and decreases in the cut–off value U due to the decay of $\mathcal{F}\mu$. The second is the stochastic error and increases in U by the growth of $\Delta\psi_n$. For growing sample size n the term $\Delta\psi_n$ becomes smaller so that the stochastic error decays even if we let $U \to \infty$ as $n \to \infty$. For $\sigma = 0$ the term $\Delta\psi_n(u)$ grows polynomially in u so that we can let U tend polynomially to infinity, whereas for $\sigma > 0$ it grows exponentially in u and we can let U tend only logarithmically to infinity. This is the reason for the polynomial and logarithmic convergence rates in the cases $\sigma = 0$ and $\sigma > 0$, respectively. For fixed sample size the cut–off value U is the crucial tuning parameter in this method and allows a trade–off between the error terms. The influence of the cut–off value U is analogous to the influence of the bandwidth h on kernel estimators, more precisely U^{-1} corresponds to h. The other estimation errors allow similar decompositions as $\Delta\hat{\sigma}^2$ in (2.29) and they are given by equations (4.1), (4.2) and (4.3) in the proof of Theorem 4.1.

To simplify the asymptotic analysis of the stochastic errors, we do not work with the regression model (2.6) but with the Gaussian white noise model. This is an idealized observation scheme, where the terms are easier to analyze. At the same time asymptotic results may be transferred to the regression model. The Gaussian white noise model is given by

$$\mathrm{d}Z_n(x) = \mathcal{O}(x)\mathrm{d}x + \epsilon_n\delta(x)\mathrm{d}W(x), \qquad x \in \mathbb{R}, \tag{2.30}$$

where W is a two–sided Brownian motion, $\delta \in L^2(\mathbb{R})$ and $\epsilon_n > 0$. In the case of equidistant design the precise connection to the regression model (2.6) is given by $\delta(x_j) = \delta_j$ and $\epsilon_n = \frac{n+1}{n-1}(x_n - x_1)n^{-1/2}$, where x_1 and x_n are the minimal and maximal design points and where we assume that the range of observations $(x_n - x_1)$ grows slower than $n^{1/2}$ such that $\epsilon_n \to 0$ as $n \to \infty$. Transferring asymptotic results from the Gaussian white noise model to the regression model is formally justified by Le Cam's concept of asymptotic equivalence, see Le Cam and Yang (2000). In particular, it can be used to transfer lower bounds and confidence statements. Brown and Low (1996) show that the regression (2.6) with Gaussian errors is asymptotically equivalent to the Gaussian white noise model (2.30). For non–Gaussian errors we refer to Grama and Nussbaum (2002). Their main assumption on the errors is slightly more than Hellinger differentiability, which is a smoothness assumption on the distributions of the errors. To be more precise on this asymptotic equivalence, we restrict the Gaussian white noise model to a sequence of growing intervals $[x_1 - \Delta_n, x_n + \Delta_n]$ with $\Delta_n := (x_n - x_1)/(n-1)$ and assume as a simplification that the observations in the regression model are equidistant with mesh size Δ_n. We suppose $\delta^2 > 0$ to be an absolutely continuous function and $|\frac{\partial}{\partial x} \log \delta(x)| \leqslant C$ to hold for some $C < \infty$. The functions $\mathcal{O}$ are uniformly bounded by $\mathcal{O}(x) = S^{-1}\mathcal{C}(x,T) - (1-e^x)^+ \leqslant 1$ and uniformly Lipschitz by $|\mathcal{O}'(x)| = |\int_{-\infty}^x \mathcal{O}''(x)\mathrm{d}x - \mathbf{1}_{\{x>0\}} + e^{(\gamma-\lambda)T}\mathbf{1}_{\{x>\gamma T, \sigma=0\}}| \leqslant 4 + e^{RT}$, where we used Proposition 2.1 in Belomestny and Reiß (2006a) and $|\gamma| \leqslant R$. These properties of $\mathcal{O}$ are used to apply Corollary 4.2 in Brown and Low (1996), which yields the asymptotic equivalence of the regression model (2.6) with Gaussian errors and the Gaussian white noise model (2.30) each restricted to the intervals $[x_1 - \Delta_n, x_n + \Delta_n]$. In Chapter 7 we will also treat nonequidistant design in the simulations and in the empirical study. For this reason we briefly mention the asymptotic equivalence for nonequidistant design. To this end, we consider both models on intervals $I_n := [\alpha_n, \beta_n]$ with $\lim_{n\to\infty} \alpha_n = -\infty$ and $\lim_{n\to\infty} \beta_n = \infty$. We assume that there are cumulative distribution functions H_n on I_n which are absolutely continuous on I_n and satisfy $H_n'(x) = h_n(x) > 0$ almost everywhere on I_n. The design points are given by $x_j = H^{-1}(j/(n+1))$, $j = 1, \dots, n$. Then the regression model (2.6) is asymptotically equivalent to the Gaussian white noise model (2.30) where $\epsilon_n \delta(x)$ is replaced by $n^{-1/2}\delta(x)h_n(x)^{-1/2}$.

The stochastic errors involve the term $\Delta\psi_n(Uu)$, which is a difference between two logarithms. We define the empirical version of $\mathcal{FO}$ by $\mathcal{F}\,\mathcal{O}_n := \mathcal{F}(\,\mathrm{d}Z_n)$. It is obtained by applying the Fourier transform directly to the Gaussian white noise model (2.30) and thus the intermediate step of constructing an empirical version of $\mathcal{O}$ may be omitted. For $z, z' \in \mathbb{C} \setminus \{0\}$ and $\kappa > 0$ it holds $\log_{\geqslant\kappa}(z) - \log(z') = \log_{\geqslant\kappa/|z'|}(z/z')$. That yields

$$\Delta\psi_n(Uu) = \frac{1}{T} \log_{\geqslant\kappa^U(u)} \Big(1 + \frac{\epsilon_n\, iUu(1+iUu)}{1 + iUu(1+iUu)\mathcal{FO}(Uu)} \int_{-\infty}^{\infty} e^{iUux}\delta(x)\mathrm{d}W(x)\Big), \tag{2.31}$$

where $\kappa^U(u) := \kappa(Uu)/|1 + iUu(1+iUu)\mathcal{FO}(Uu)| \leqslant 1/2$, see (Belomestny and Reiß, 2006a, (6.3)). We define a linearization $\mathcal{L}_{n,U}$ of the logarithm and the remainder term $\mathcal{R}_{n,U}$ by

$$\mathcal{L}_{n,U}(u) := \frac{\epsilon_n\, iUu(1+iUu)}{T(1 + iUu(1+iUu)\mathcal{FO}(Uu))} \int_{-\infty}^{\infty} e^{iUux}\delta(x)\mathrm{d}W(x), \tag{2.32}$$

$$\mathcal{R}_{n,U}(u) := \Delta\psi_n(Uu) - \mathcal{L}_{n,U}(u). \tag{2.33}$$

This linearization will be an important step in proving asymptotic normality in Chapter 4. We will see that on appropriate conditions the linearization $\mathcal{L}_{n,U}$ determines the asymptotic distribution of the errors and that the error caused by the remainder term $\mathcal{R}_{n,U}$ is asymptotically negligible. $\mathcal{L}_{n,U}$ is a Gaussian process and we will devote the next chapter to the analysis of this Gaussian process.

3 On a related Gaussian process

Applying the estimation method to observations in the Gaussian white noise model, leads naturally to the Gaussian process $\mathcal{L}_{n,U}$, which we will study in this chapter. Continuity and boundedness are classical topics in the theory of Gaussian processes and we will investigate these properties for $\mathcal{L}_{n,U}$. In addition, we will treat hitting probabilities of $\mathcal{L}_{n,U}$, which are motivated by the use of the distinguished logarithm in the definitions of ψ and ψ_n in (2.10) and (2.11), respectively, as well as by equality (2.31) for $\Delta\psi_n$. The distinguished logarithm of a function $\varphi : [-U, U] \to \mathbb{C}$ with $\varphi(0) = 1$ exists if it is continuous and does not vanish (Chung, 1974, Thm. 7.6.2). The estimators are based on ψ_n and thus implicitly rely on the existence of the distinguished logarithm. Nevertheless, for the estimators $\hat{\sigma}^2$, $\hat{\lambda}$ and $\hat{\mu}(0)$ the distinguished logarithm may be avoided by using the identity $\mathrm{Re}(\log(z)) = \log(|z|)$, so that it suffices to use the usual logarithm of the positive real numbers. But for the estimators $\hat{\gamma}$ and $\hat{\mu}(x)$, $x \neq 0$, the imaginary part of ψ_n will in general contribute to the estimators so that the use of the distinguished logarithm is essential. The distinguished logarithm in the definition of ψ in (2.10) is well–defined since $\varphi_T(u - i)$ is continuous and does not vanish for all $u \in \mathbb{R}$, which can be seen from the Lévy–Khintchine representation. By $\psi_n = \psi + \Delta\psi_n$ we conclude that ψ_n being well–defined is equivalent to $\Delta\psi_n$ being well–defined. The distinguished logarithm in the definition of $\Delta\psi_n(Uu)$ in (2.31) is well–defined for $u \in [-1, 1]$ whenever the argument of the logarithm $1 + T\mathcal{L}_{n,U}$ is continuous and $1 + T\mathcal{L}_{n,U}(u) \neq 0$ for all $u \in [-1, 1]$. Thus we are interested in conditions ensuring that $\mathcal{L}_{n,U}$ is continuous and that

$$\mathcal{L}_{n,U}(u) \neq -\frac{1}{T} \quad \text{for all } u \in [-1, 1] \tag{3.1}$$

either with high probability or almost surely.

There is another property of $\mathcal{L}_{n,U}$ that we would like to study. The proofs of the asymptotic normality results for the estimators rely on the approximation of $\Delta\psi_n(Uu)$ by $\mathcal{L}_{n,U}(u)$. This is a good approximation if the argument of the logarithm in the definition of $\Delta\psi_n$ is close to one, which is equivalent to $\mathcal{L}_{n,U}$ being small. So by bounding $\mathcal{L}_{n,U}$, we can bound the remainder $\mathcal{R}_{n,U}$. Since we are integrating over the unit interval in the decompositions of the estimation errors (2.29)–(4.3), we will bound $\mathcal{L}_{n,U}$ uniformly on the unit interval.

This chapter is divided into two parts. In Section 3.1 we study continuity and boundedness of $\mathcal{L}_{n,U}$ and in Section 3.2 we proof a general result on hitting probabilities and apply it to our situation. This chapter includes the results by Söhl (2010), where most of the material can be found.

3.1 Continuity and boundedness

We first show an auxiliary lemma, where we assume

Condition 3.1. There is a $p > 0$ such that $\int_{-\infty}^{\infty}(1+|x|)^p \delta(x)^2 \mathrm{d}x < \infty$.

Lemma 3.2. *Let δ fulfill Condition 3.1. Then there exists a number $c > 0$ such that the stochastic process $X(v) = \int_{-\infty}^{\infty} e^{ivx}\delta(x)\,\mathrm{d}W(x)$ satisfies for all $u, v \in \mathbb{R}$*

$$\sqrt{\mathbb{E}[|X(u) - X(v)|^2]} \leqslant c|u-v|^{\min(p/2,1)}.$$

Proof. Condition 3.1 is satisfied for $q := \min(p, 2)$ as well. Without loss of generality we assume $u \neq v$ and conclude that

$$\begin{aligned}
\mathbb{E}\left[|X(u)-X(v)|^2\right] &= \mathbb{E}\left[\left|\int_{-\infty}^{\infty}(e^{iux}-e^{ivx})\delta(x)\mathrm{d}W(x)\right|^2\right] \\
&= \int_{-\infty}^{\infty} |e^{iux}-e^{ivx}|^2\delta(x)^2\mathrm{d}x \\
&\leqslant \int_{-\infty}^{\infty} \min(4,(u-v)^2x^2)\delta(x)^2\mathrm{d}x \\
&= \int_{|x|\geqslant 2|u-v|^{-1}} 4\delta(x)^2\mathrm{d}x + \int_{|x|<2|u-v|^{-1}} (u-v)^2x^2\delta(x)^2\mathrm{d}x \\
&\leqslant \int_{|x|\geqslant 2|u-v|^{-1}} 4\left(\frac{|x|}{2|u-v|^{-1}}\right)^q \delta(x)^2\mathrm{d}x \\
&\quad + \int_{|x|<2|u-v|^{-1}} \left(\frac{2|u-v|^{-1}}{|x|}\right)^{2-q} (u-v)^2x^2\delta(x)^2\mathrm{d}x \\
&= 2^{2-q}|u-v|^q \int_{-\infty}^{\infty} |x|^q\delta(x)^2\mathrm{d}x.
\end{aligned}$$

This shows the lemma. $\square$

The next proposition shows that on Condition 3.1 the Gaussian process $\mathcal{L}_{n,U}$ is continuous and bounded while also giving a bound for the expected value of the supremum. It is Proposition 1 by Söhl (2012) extended by a continuity statement. We shall use the Landau notation $A(x) = O(B(x))$ as $x \to \infty$, meaning that there exist $M > 0$ and $x_0 \in \mathbb{R}$ such that $A(x) \leqslant MB(x)$ for all $x \geqslant x_0$.

Proposition 3.3. *Grant Condition 3.1. $\mathcal{L}_{n,U}$ has a version which is almost surely continuous on the whole real line. Moreover, if for all $U > 0$ the processes $\mathcal{L}_{n,U}$ are almost surely continuous on $[-1,1]$, then for each $q \geqslant 1$*

$$\mathbb{E}\left[\sup_{u\in[-1,1]}|\mathcal{L}_{n,U}(u)|^q\right]^{1/q} = \begin{cases} O(\epsilon_n U^2\sqrt{\log(U)}), & \textit{for } \sigma = 0, \\ O(\epsilon_n U^2 \exp(T\sigma^2U^2/2)), & \textit{for } \sigma > 0, \end{cases} \quad \textit{as } U \to \infty.$$

Proof. First we define $X(u) := \int_{-\infty}^{\infty} e^{iux}\delta(x)\mathrm{d}W(x)$. Since $X(-u) = \overline{X(u)}$ it suffices to consider suprema of the absolute value $|X(u)|$ over positive index sets. We assumed that there is an $p > 0$ such that $\int_{-\infty}^{\infty}(1 + |x|)^p\delta(x)^2\mathrm{d}x < \infty$. Lemma 3.2 shows that there exists a number $c > 0$ such that $\sqrt{\mathbb{E}[|X(u) - X(v)|^2]} \leqslant c|u - v|^H$ for all $u, v \in \mathbb{R}$ with $H := \min(p/2, 1) \in (0, 1]$. Denote by $N_\rho(I, r)$ the covering number, that is the minimum number of closed balls of radius r in the metric ρ with centers in I that cover I. We define

$$\rho(u, v) := c|u - v|^H$$

and

$$d(u, v) := \sqrt{\mathbb{E}[|X(u) - X(v)|^2]}.$$

By the above inequality $d(u, v) \leqslant \rho(u, v)$ for all $u, v \in \mathbb{R}$. A ball of radius r in the metric ρ covers an interval of length $2(r/c)^{1/H}$. Thus, it holds

$$N_\rho([0, U], r) = \left\lceil U\,(c/r)^{1/H}/2 \right\rceil,$$

where $\lceil a \rceil$ is the smallest integer equal or larger than a. The radius of the smallest ball with center in $[0, U]$ that contains $[0, U]$ is $c(U/2)^H$ with respect to the metric ρ. There exists $D < \infty$ such that $d(u, v) \leqslant D$ for all $u, v \in \mathbb{R}$. For U large enough such that $U \geqslant (D/c)^{1/H}$ we have the entropy bound

$$\begin{aligned} J([0, U], d) &:= \int_0^\infty (\log(N_d([0, U], r)))^{1/2}\,\mathrm{d}r = \int_0^D (\log(N_d([0, U], r)))^{1/2}\,\mathrm{d}r \\ &\leqslant \int_0^D (\log(N_\rho([0, U], r)))^{1/2}\,\mathrm{d}r \leqslant \int_0^D \left(\log\left(U\,(c/r)^{1/H}\right)\right)^{1/2}\mathrm{d}r \qquad (3.2) \\ &\leqslant H^{-1/2}\int_0^D \left(\log\left(U^H c/r\right)\right)^{1/2}\mathrm{d}r, \end{aligned}$$

here we substitute $r = U^H cs$,

$$\leqslant H^{-1/2}U^H c\int_0^{D/(U^H c)} (\log(1/s))^{1/2}\,\mathrm{d}s. \qquad (3.3)$$

This integral is solved by

$$\int_0^x \sqrt{\log y^{-1}}\mathrm{d}y = \frac{\sqrt{\pi}}{2} - \frac{\sqrt{\pi}}{2}\mathrm{Erf}(\sqrt{\log x^{-1}}) + x\sqrt{\log x^{-1}},$$

where $\mathrm{Erf}(y) = \frac{2}{\sqrt{\pi}}\int_0^y e^{-t^2}\mathrm{d}t$. For all $y > 0$ the estimate $1 - \mathrm{Erf}(y) \leqslant \exp(-y^2)/(\sqrt{\pi}y)$ holds, which is a standard estimate for the c.d.f. of the Gaussian distribution, see Lemma 22.2 in Klenke (2007).

For each $H > 0$ this yields $\widetilde{c} > 0$ such that for all $x \in (0, 1/2^H)$

$$\int_0^x \sqrt{\log y^{-1}} \mathrm{d}y \leqslant \widetilde{c} x \sqrt{\log x^{-1}}.$$

Thus, (3.3) can be bounded by

$$(H^{-1/2} \widetilde{c} D) \sqrt{\log(U^H c/D)} = O((\log U)^{1/2})$$

as $U \to \infty$. Consequently, $(\log U)^{1/2}$ is an asymptotic upper bound of the entropy integral (3.2). By Dudley's theorem (e.g., see Kahane, 1985, p. 219) there is for all $U > 0$ a version of X which is almost surely continuous on $[-U, U]$ with respect to d and by Lemma 3.2 also with respect to the Euclidean metric. Two versions of the same stochastic process are indistinguishable if they are both almost surely continuous and if the index set is an interval (Klenke, 2007, Lem. 21.5). Thus, there is a version of X and subsequently of $\mathcal{L}_{n,U}$ which is almost surely continuous with respect to the Euclidean metric on $\mathbb{R}$.

Let X' be an almost surely continuous version of X. Then

$$\frac{\epsilon_n i U u(1 + iUu)}{1 + iUu(1 + iUu)\,\mathcal{F}\,\mathcal{O}(Uu)} X'(Uu) \tag{3.4}$$

is an almost surely continuous version of $\mathcal{L}_{n,U}$. If $\mathcal{L}_{n,U}$ is almost surely continuous on $[-1, 1]$ then $\mathcal{L}_{n,U}$ and the process (3.4) are indistinguishable on $[-1, 1]$.

Dudley's theorem yields for all $q \geqslant 1$

$$\mathbb{E}\left[\sup_{u \in [-U,U]} |\operatorname{Re}(X'(u))|^q\right] = O((\log U)^{q/2}) \tag{3.5}$$

and

$$\mathbb{E}\left[\sup_{u \in [-U,U]} |\operatorname{Im}(X'(u))|^q\right] = O((\log U)^{q/2})$$

as $U \to \infty$. We estimate from above for all $q \geqslant 1$

$$\begin{aligned}
&\mathbb{E}\left[\sup_{u \in [-1,1]} |\mathcal{L}_{n,U}(u)|^q\right] \\
&\leqslant \sup_{u \in [-U,U]} \left|\frac{\epsilon_n i u(1 + iu)}{T(1 + iu(1 + iu)\mathcal{F}\mathcal{O}(u))}\right|^q \mathbb{E}\left[\sup_{u \in [0,U]} |X'(u)|^q\right] \\
&\leqslant \left(\frac{\epsilon_n U \sqrt{1 + U^2}}{T \exp(T(-\sigma^2 U^2/2 + \sigma^2/2 + \gamma - \lambda - \|\mathcal{F}\mu\|_\infty))}\right)^q \mathbb{E}\left[\sup_{u \in [0,U]} |X'(u)|^q\right] \\
&= O\left(\left(\epsilon_n U^2 \sqrt{\log(U)} \exp(T\sigma^2 U^2/2)\right)^q\right)
\end{aligned} \tag{3.6}$$

as $U \to \infty$. This completes the proof for the case $\sigma = 0$. For $\sigma > 0$ we observe

$$\mathbb{E}\left[\sup_{u\in[-1,1]} |\mathcal{L}_{n,U}(u)|^q\right] \leqslant \mathbb{E}\left[\sup_{|u|\leqslant U-1} |\mathcal{L}_{n,1}(u)|^q\right] + \mathbb{E}\left[\sup_{|u|\in[U-1,U]} |\mathcal{L}_{n,1}(u)|^q\right].$$

By the previous considerations the growth of the first part can be bounded by

$$\left(\epsilon_n(U-1)^2\sqrt{\log(U-1)}\exp(T\sigma^2(U-1)^2/2)\right)^q = O\left(\left(\epsilon_n U^2\exp(T\sigma^2U^2/2)\right)^q\right). \quad (3.7)$$

For the second part we note that as in (3.2) we have

$$J([U-1,U],d) \leqslant \int_0^D (\log(N_\rho([U-1,U],r)))^{1/2}\,\mathrm{d}r = \int_0^D (\log(N_\rho([0,1],r)))^{1/2}\,\mathrm{d}r$$

and thus the entropy does not depend on U. For $u \in [U-1,U]$ the process X' does not contribute a logarithmic factor and it holds

$$\mathbb{E}\left[\sup_{u\in[-1,1]} |\mathcal{L}_{n,U}(u)|^q\right] = O\left(\left(\epsilon_n U^2\exp(T\sigma^2U^2/2)\right)^q\right)$$

as $U \to \infty$. □

Proposition 3.3 yields a bound for the expected value of the supremum of $\mathcal{L}_{n,U}$ on $[-1,1]$. This is important in order to control the remainder $\mathcal{R}_{n,U}$ when we approximate $\Delta\psi_n$ by $\mathcal{L}_{n,U}$ and will be used to prove asymptotic normality in Chapter 4.

3.2 Hitting probabilities

Let us now discuss conditions on which the distinguished logarithm used in the estimation method is well–defined. We assume Condition 3.1 to be fulfilled. By Proposition 3.3 there is an almost surely continuous version of $\mathcal{L}_{n,U}$ and in the following we will always assume $\mathcal{L}_{n,U}$ to be such a version. For the cut–off value we assume that it is chosen to satisfy $U_n \to \infty$ and either $\epsilon_n U_n^2\sqrt{\log(U_n)} \to 0$ or $\epsilon_n U_n^2\exp(T\sigma^2U_n^2/2) \to 0$ as $n \to \infty$ depending on whether $\sigma = 0$ or $\sigma > 0$. The respective expressions appear in the bound on $\mathbb{E}\left[\sup_{u\in[-1,1]} |\mathcal{L}_{n,U_n}(u)|\right]$ in Proposition 3.3. By Markov's inequality we conclude $\mathbb{P}(\sup_{u\in[-1,1]} |\mathcal{L}_{n,U_n}(u)| \geqslant 1/T) = O(\epsilon_n U_n^2\sqrt{\log(U_n)})$ as $n \to \infty$ for $\sigma = 0$ and $\mathbb{P}(\sup_{u\in[-1,1]} |\mathcal{L}_{n,U_n}(u)| \geqslant 1/T) = O(\epsilon_n U_n^2\exp(T\sigma^2U_n^2/2))$ as $n \to \infty$ for $\sigma > 0$ and thus the probability that $\mathcal{L}_{n,U}(u) \neq -1/T$ for all $u \in [-1,1]$ as required in (3.1) tends to one. Consequently, the probability of the sets where the estimators are possibly not well–defined tends to zero for $n \to \infty$. This result is very similar to Theorem 2.3 by Gugushvili (2009) for Lévy processes observed at low frequency. However, we will see that ψ_n is almost surely well–defined by (2.11) on a slightly stronger assumption on δ and this implies that the estimators are even almost surely well–defined.

To this end, we will show that $1 + iu(1 + iu)\,\mathcal{F}\,\mathcal{O}_n(u)$ is almost surely continuous and does not hit zero almost surely. Continuity of $\mathcal{L}_{n,U}$ is equivalent to continuity of $1+iu(1+iu)\,\mathcal{F}\,\mathcal{O}_n(u)$. So the main difficulty is to proof that zero is *polar* meaning that the Gaussian process $1 + iu(1 + iu)\,\mathcal{F}\,\mathcal{O}_n(u)$ does not hit zero almost surely. Hitting probabilities and polar sets have been studied for Gaussian processes on the assumption that the components of the Gaussian process consists of independent copies of the same Gaussian process. Identifying $\mathbb{C}$ with $\mathbb{R}^2$ we see that $\mathcal{F}\,\mathcal{O}_n$ is a Gaussian process taking values in $\mathbb{R}^2$. But the components are in general not independent copies of the same Gaussian process. So we will study hitting probabilities and polar sets for Gaussian processes where the components are not independent copies of the same process.

More generally, we will consider Gaussian random fields, which are generalizations of Gaussian processes to multidimensional index sets. Let $X = \{X(t) | t \in I \subseteq \mathbb{R}^N\}$ be a centered Gaussian random field with values in $\mathbb{R}^d$, where I is bounded. We will call X an (N, d)–*Gaussian random field.* The *intrinsic covariance metric* also called canonical metric associated with the Gaussian random field is $\sqrt{\mathbb{E}\left[\|X(s) - X(t)\|^2\right]}$, where $\|\bullet\|$ denotes the Euclidean metric. Polar sets for Gaussian random fields are investigated in Weber (1983) under the assumptions that the components are independent copies of the same random field, that the variance is constant and that $\sqrt{\mathbb{E}\left[\|X(s) - X(t)\|^2\right]} \leqslant c\|s - t\|^\beta$ holds with constants $c, \beta > 0$. The recent works Xiao (2009) and Biermé et al. (2009) consider the anisotropic metric

$$\rho(s,t) := \sum_{j=1}^{N} |s_j - t_j|^{H_j} \tag{3.8}$$

with $H \in (0,1]^N$ and assume $\sqrt{\mathbb{E}\left[\|X(s) - X(t)\|^2\right]} \leqslant c\rho(s,t)$. In addition they require the variance only to be bounded from below. We substitute the assumptions on the variance and on the independent copies in the components by the milder assumption that the eigenvalues of the covariance matrix are bounded from below. The random fields in the components neither need to be identically distributed nor independent. Hence, we require weaker assumptions on the dependency structure of the components of the Gaussian random field than Weber (1983), Xiao (2009) and Biermé et al. (2009). It follows from an upper bound on the hitting probabilities of X that sets with Hausdorff dimension smaller than $d - \sum_{j=1}^N 1/H_j$ are polar. Our results allow for a translation of the Gaussian random field X by a random field, that is independent of X and whose sample functions are Lipschitz continuous with respect to the metric ρ.

In Section 3.2.1 we will proof a theorem on hitting probabilities for Gaussian random fields. In Section 3.2.2 we will apply this theorem to the Gaussian process $1 + iu(1 + iu)\,\mathcal{F}\,\mathcal{O}_n(u)$ and we will conclude that ψ_n is almost surely well–defined. Section 3.2.3 contains the proof of Lemma 3.6.

3.2.1 General results

Let X be an (N,d)–Gaussian random field. Recall that we suppose the index set I to be bounded. We will assume the following two conditions.

Condition 3.4. There is a constant $c > 0$ such that we have $\sqrt{\mathbb{E}\left[\|X(s)-X(t)\|^2\right]} \leqslant c\rho(s,t)$ for all $s,t \in I$.

Condition 3.5. There is a constant $\lambda > 0$ such that for all $t \in I$ and for all $e \in \mathbb{R}^d$ with $\|e\| = 1$ we have $\mathbb{E}[(\sum_{j=1}^{d} e_j X_j(t))^2] \geqslant \lambda$.

Condition 3.4 bounds the intrinsic covariance metric in terms of the anisotropic metric ρ. Condition 3.5 bounds the eigenvalues of the covariance matrix from below. It excludes, for example, cases where X takes values only in some vector subspace.

We will use a uniform modulus of continuity, see (69) in (Xiao, 2009, p. 167). We restate this result in the next inequality. Let X be an (N,d)–Gaussian random field, that satisfies Condition 3.4. Then there is a version X' of X and a constant $\widetilde{c} > 0$ such that almost surely the following inequality holds:

$$\limsup_{\varepsilon \downarrow 0} \sup_{s,t \in I, \rho(s,t) \leqslant \varepsilon} \frac{\|X'(s) - X'(t)\|}{\varepsilon\sqrt{\log(\varepsilon^{-1})}} \leqslant \widetilde{c}. \tag{3.9}$$

We will always assume that X is a version, which satisfies (3.9). Let

$$\mathrm{Lip}_\rho(L) := \{f : I \to \mathbb{R}^d \mid \|f(s) - f(t)\| \leqslant L\rho(s,t)\ \forall s,t \in I\}$$

denote the *L–Lipschitz functions* with respect to the metric ρ. In each direction j the functions in $\mathrm{Lip}_\rho(L)$ are Hölder continuous with exponent H_j. We denote by $B_\rho(t,r) := \{s \in \mathbb{R}^N | \rho(s,t) \leqslant r\}$ the closed ball of radius r around t. For the theorem on hitting probabilities we will use the following lemma, which is proved in Section 3.2.3.

Lemma 3.6. *Let X be an (N,d)–Gaussian random field, that satisfies Conditions 3.4 and 3.5. Then for each $L \geqslant 0$ there is a constant $C > 0$ such that for all $t \in I$, for all $r > 0$ and for all functions $f \in Lip_\rho(L)$ we have*

$$\mathbb{P}\left(\inf_{s \in B_\rho(t,r) \cap I} \|X(s) - f(s)\| \leqslant r\right) \leqslant Cr^d. \tag{3.10}$$

In the following we will recall the definitions of *Hausdorff measure* and *Hausdorff dimension* as given in Kahane (1985, p. 129). Let $A \subseteq \mathbb{R}^d$, $\alpha \in (0,d]$ and $\varepsilon > 0$. We define $\mathcal{H}_\alpha^\varepsilon(A) \in [0,\infty]$ by

$$\mathcal{H}_\alpha^\varepsilon(A) := \inf \sum_n (\operatorname{diam} B_n)^\alpha,$$

where the infimum is taken over all sets of closed balls B_n with diameter $\operatorname{diam} B_n$ less or equal to ε such that there union covers A. As $\varepsilon \to 0$ the numbers $\mathcal{H}_\alpha^\varepsilon(A)$ increase and we denote the limit by $\mathcal{H}_\alpha(A) \in [0,\infty]$, and call $\mathcal{H}_\alpha(A)$ the α–dimensional Hausdorff measure of A. For $0 < \alpha < \beta \leqslant d$ we have

$$\sum_n (\operatorname{diam} B_n)^\beta \leqslant \sup_n (\operatorname{diam} B_n)^{\beta-\alpha} \sum_n (\operatorname{diam} B_n)^\alpha,$$

such that $\mathcal{H}_\alpha(A) < \infty$ implies $\mathcal{H}_\beta(A) = 0$ and $\mathcal{H}_\beta(A) > 0$ implies $\mathcal{H}_\alpha(A) = \infty$. We obtain $\sup\{\alpha | \mathcal{H}_\alpha(A) = \infty\} = \inf\{\beta | \mathcal{H}_\beta(A) = 0\}$ and call this number the Hausdorff dimension of A. Recall that $Q = \sum_{j=1}^N 1/H_j$ with H_j as in the definition of the metric ρ.

Theorem 3.7. *Let X be an (N,d)–Gaussian random field that satisfies Conditions 3.4 and 3.5. If $Q < d$, then for each $L \geqslant 0$ there is a constant $C > 0$ such that all Borel sets $F \subseteq \mathbb{R}^d$ and all random fields Y which are independent of X and whose sample functions are all in $Lip_\rho(L)$ satisfy*

$$\mathbb{P}\left(\exists s \in I : X(s) + Y(s) \in F\right) \leqslant C\mathcal{H}_{d-Q}(F). \tag{3.11}$$

Proof. By Fubini's theorem it suffices to show for all functions $f \in \mathrm{Lip}_\rho(L)$

$$\mathbb{P}\left(\exists s \in I : X(s) + f(s) \in F\right) \leqslant C\mathcal{H}_{d-Q}(F).$$

We choose some constant $\gamma > \mathcal{H}_{d-Q}(F)$. By definition of the Hausdorff measure there is a set of balls $\{B(x_l, r_l) : l = 0, 1, 2, \dots\}$ such that

$$F \subseteq \bigcup_{l=0}^{\infty} B(x_l, r_l) \quad \text{and} \quad \sum_{l=0}^{\infty} (2r_l)^{d-Q} \leqslant \gamma. \tag{3.12}$$

For all j we cut the bounded index set I orthogonal to the j–axis with distance $(r_l/N)^{1/H_j}$ between the cuts. Each piece of I can be covered by a single ball of radius r_l in the metric ρ. Hence there is a constant $c_8 > 0$ such that I can be covered by at most $c_8 r_l^{-Q}$ balls. We apply Lemma 3.6 to these balls. By summing up we obtain

$$\mathbb{P}\left(\exists s \in I : X(s) + f(s) \in B(x_l, r_l)\right) \leqslant c_9 r_l^{d-Q}. \tag{3.13}$$

By (3.12) and (3.13) we have

$$\begin{aligned}&\mathbb{P}\left(\exists s \in I : X(s) + f(s) \in F\right)\\ &\leqslant \sum_{l=0}^{\infty} \mathbb{P}\left(\exists s \in I : X(s) + f(s) \in B(x_l, r_l)\right) \leqslant c_{10}\gamma.\end{aligned}$$

We have $\mathbb{P}\left(\exists s \in I : X(s) + f(s) \in F\right) \leqslant c_{10}\mathcal{H}_{d-Q}(F)$, since $\gamma > \mathcal{H}_{d-Q}(F)$ was chosen arbitrarily. $\square$

3.2.2 Application

In this section, we show that ψ_n is almost surely well–defined by applying Theorem 3.7 to the Gaussian process $1 + iu(1 + iu)\,\mathcal{F}\,\mathcal{O}_n(u)$. As discussed before the definition of ψ in (2.10) we assume that the second moment of the asset price is finite such that $\mathcal{O}(x) \leqslant Ce^{-|x|}$ for some $C > 0$. Especially $x\mathcal{O}(x)$ is integrable. We require the following condition on δ, which is a stronger version of Condition 3.1.

Condition 3.8. There is a $p > 1$ such that $\int_{-\infty}^{\infty}(1 + |x|)^p\delta(x)^2\mathrm{d}x < \infty$.

For example, if $\delta \in L^2(\mathbb{R})$ and $\delta(x) = O(|x|^{-p})$ for $|x| \to \infty$ with $p > 1$, then the condition is satisfied. Condition 3.8 (or the weaker Condition 3.1) and Lemma 3.2 imply the uniform modulus of continuity (3.9) for a version of $X(v) := \int_{-\infty}^{\infty} e^{ivx}\delta(x)\,\mathrm{d}W(x)$. We will assume that X is a version that satisfies (3.9). Thus in the definition of ψ_n in (2.11) the argument of the logarithm is almost surely continuous.

Lemma 3.9. *Let δ fulfill Condition 3.8. Then ψ_n is almost surely well–defined.*

Proof. We have to show that almost surely the argument of the logarithm does not hit zero. The process $1 + iv(1 + iv)\mathcal{F}\mathcal{O}_n(v)$ equals 1 at $v = 0$. It suffices to consider the process on $\mathbb{R}\backslash\{0\}$. We rewrite the process as

$$iv(1+iv)\left(\frac{1}{iv(1+iv)} + \mathcal{F}\mathcal{O}(v) + \epsilon_n \int_{-\infty}^{\infty} e^{ivx}\delta(x)\,\mathrm{d}W(x)\right).$$

We define

$$f(v) := \frac{1}{iv(1+iv)} + \mathcal{F}\mathcal{O}(v) \quad \text{and} \quad X(v) := \epsilon_n \int_{-\infty}^{\infty} e^{ivx}\delta(x)\,\mathrm{d}W(x).$$

We identify $\mathbb{C}$ with $\mathbb{R}^2$. X is a Gaussian process that takes values in $\mathbb{R}^2$. If we restrict X to a bounded index set, then X is an (1,2)–Gaussian random field. We will apply Theorem 3.7 to X, $Y = f$ and $F = \{0\}$. By Lemma 3.2 there is a constant $c > 0$ such that for all $u, v \in \mathbb{R}$ the inequality

$$\sqrt{\mathbb{E}[\|X(u) - X(v)\|^2]} \leqslant c|u - v|^{\min(p/2,1)}. \tag{3.14}$$

holds. This gives reason to the definition $\rho(u, v) := |u - v|^H$ with $H = \min(p/2, 1) \in (1/2, 1]$. Thus Condition 3.4 is satisfied and we have $d - Q = 2 - 1/H > 0$.

It remains to show that Condition 3.5 is fulfilled and that f is Lipschitz continuous with respect to the metric ρ. For $\delta = 0 \in L^2(\mathbb{R})$ we have $\psi_n = \psi$ and thus ψ_n is well–defined. We will now show that the covariance matrix of $X(v)$ is not degenerated if $\delta \neq 0 \in L^2(\mathbb{R})$ and $v \neq 0$. Let $e \in \mathbb{R}^2$ such that $e_1^2 + e_2^2 = 1$. Then there is $\varphi \in [0, 2\pi]$ such that $e_1 = \sin\varphi$ and $e_2 = \cos\varphi$. Consider X as a $\mathbb{R}^2$–valued stochastic process. The Itô isometry yields

$$\mathbb{E}[(e_1X_1(v) + e_2X_2(v))^2] = \mathbb{E}\left[\left(\epsilon_n \int_{-\infty}^{\infty}(e_1\cos(vx) + e_2\sin(vx))\delta(x)\,\mathrm{d}W(x)\right)^2\right]$$

$$= \epsilon_n^2 \int_{-\infty}^{\infty} (e_1 \cos(vx) + e_2 \sin(vx))^2 \delta(x)^2 \, dx$$
$$= \epsilon_n^2 \int_{-\infty}^{\infty} (\sin(\varphi + vx))^2 \delta(x)^2 \, dx > 0.$$

The function

$$\mathbb{R} \times [0, 2\pi] \to \mathbb{R}, (v, \varphi) \mapsto \epsilon_n^2 \int_{-\infty}^{\infty} (\sin(\varphi + vx))^2 \delta(x)^2 dx$$

is continuous by dominated convergence. On $([-V, -1/V] \cup [1/V, V]) \times [0, 2\pi]$ it takes a minimum $\lambda_V > 0$ for $V > 0$. Hence Condition 3.5 is fulfilled on the index set $I_V = [-V, -1/V] \cup [1/V, V]$.

Since $x\mathcal{O}(x)$ is integrable we have that $\mathcal{F}\mathcal{O}$ is Lipschitz continuous on $\mathbb{R}$. $1/(iv(1+iv))$ is Lipschitz continuous on sets bounded away from zero. Hence f is Lipschitz continuous on I_V. Since I_V is bounded it follows that f is Lipschitz continuous with respect to the metric ρ on I_V.

Thus we may apply Theorem 3.7 to the index set $I_V = [-V, -1/V] \cup [1/V, V]$. Since $\mathcal{H}_{d-Q}(\{0\}) = 0$ we obtain $\mathbb{P}(\exists v \in I_V : X(v) + f(v) = 0) = 0$. Because $V > 0$ was chosen arbitrarily the lemma follows. □

3.2.3 Proof of Lemma 3.6

For all integers $n \geqslant 1$ we define $\varepsilon_n := r \exp(-2^{n+1})$ and denote by $N_n := N_\rho(B_\rho(t, r) \cap I, \varepsilon_n)$ the covering number, that is the minimum number of ρ–balls with radii ε_n and centers in $B_\rho(t, r) \cap I$ that are needed to cover $B_\rho(t, r) \cap I$. We have the inclusion $B_\rho(t, r) \subseteq \prod_{j=1}^N [t_j - r^{1/H_j}, t_j + r^{1/H_j}]$. On the other hand each set $\prod_{j=1}^N [s_j, s_j + (\varepsilon_n/N)^{1/H_j}]$ can be covered by a single ball with radius ε_n. Hence there is a constant $c_1 > 0$ independent of n such that $N_n \leqslant \prod_{j=1}^N ((2rN/\varepsilon_n)^{(1/H_j)} + 1) \leqslant c_1 \exp(Q 2^{n+1})$ where $Q = \sum_{j=1}^N 1/H_j$.

We denote by $\{t_i^{(n)} \in B_\rho(t, r) \cap I | 1 \leqslant i \leqslant N_n\}$ a set of points such that the balls with the centers $\{t_i^{(n)}\}$ and radii ε_n cover $B_\rho(t, r) \cap I$. We define

$$r_n := \beta \varepsilon_n 2^{\frac{n+1}{2}},$$

where $\beta > \widetilde{c}$ is some constant to be determined later. For all integers $n, k \geqslant 1$ and $1 \leqslant i \leqslant N_k$, we define the following events

$$A_i^{(k)} := \Big\{ \|X(t_i^{(k)}) - f(t_i^{(k)})\| \leqslant r + \sum_{l=k}^{\infty} r_l \Big\}, \tag{3.15}$$

$$A^{(n)} := \bigcup_{k=1}^{n} \bigcup_{i=1}^{N_k} A_i^{(k)} = A^{(n-1)} \cup \bigcup_{i=1}^{N_n} A_i^{(n)}, \tag{3.16}$$

where the last equality only holds for $n \geqslant 2$. We will show that the probability in (3.10) can be dominated by the limit of the probabilities of the sets $A^{(n)}$

$$\mathbb{P}\left(\inf_{s\in B_\rho(t,r)\cap I}\|X(s)-f(s)\|\leqslant r\right)\leqslant\lim_{n\to\infty}\mathbb{P}(A^{(n)}). \tag{3.17}$$

For all $s\in B_\rho(t,r)\cap I$ and all $n\geqslant 1$ there exists i_n such that $\rho(s,t^{(n)}_{i_n})\leqslant\varepsilon_n$. By (3.9) we obtain almost surely

$$\limsup_{n\to\infty}\sup_{s\in I}\frac{\|X(s)-X(t^{(n)}_{i_n})\|}{r_n}\leqslant\frac{\widetilde{c}}{\beta}<1,$$

where the supremum over s is to be understood such that i_n varies according to s. Let $\kappa\in\ (\widetilde{c}/\beta,1)$. Especially there is N such that for all $n\geqslant N$ we have

$$\sup_{s\in I}\frac{\|X(s)-X(t^{(n)}_{i_n})\|}{r_n}\leqslant\kappa. \tag{3.18}$$

By going over to a possibly greater constant N, we ensure that $(1-\kappa)\widetilde{c}2^{\frac{N+1}{2}}\geqslant L$. On the event $\inf_{s\in B_\rho(t,r)\cap I}\|X(s)-f(s)\|\leqslant r$ there exists $s_0\in B_\rho(t,r)\cap I$ such that

$$\|X(s_0)-f(s_0)\|\leqslant r+\sum_{l=N+1}^{\infty}r_l. \tag{3.19}$$

Choose i_N such that $\rho(s_0,t^{(N)}_{i_N})\leqslant\varepsilon_N$. Using (3.18), (3.19) and the Lipschitz continuity of f we obtain

$$\begin{aligned}\|X(t^{(N)}_{i_N})-f(t^{(N)}_{i_N})\|&\leqslant\|X(t^{(N)}_{i_N})-X(s_0)\|+\|X(s_0)-f(s_0)\|+\|f(s_0)-f(t^{(N)}_{i_N})\|\\&\leqslant\kappa r_N+r+\sum_{l=N+1}^{\infty}r_l+L\rho(s_0,t^{(N)}_{i_N})\\&\leqslant\kappa r_N+r+\sum_{l=N+1}^{\infty}r_l+(1-\kappa)\widetilde{c}2^{\frac{N+1}{2}}\varepsilon_N\leqslant r+\sum_{l=N}^{\infty}r_l\end{aligned}$$

and (3.17) is established.

Trivially we have for $n\geqslant 2$

$$\mathbb{P}(A^{(n)})\leqslant\mathbb{P}(A^{(n-1)})+\mathbb{P}(A^{(n)}\backslash A^{(n-1)})$$

and by (3.16) we have

$$\mathbb{P}(A^{(n)}\backslash A^{(n-1)})\leqslant\sum_{i=1}^{N_n}\mathbb{P}(A^{(n)}_i\backslash A^{(n-1)}_{i'}),$$

where i' is chosen such that $\rho(t_i^{(n)}, t_{i'}^{(n-1)}) < \varepsilon_{n-1}$. We note that for $n \geqslant 2$

$$
\begin{aligned}
&\mathbb{P}(A_i^{(n)} \backslash A_{i'}^{(n-1)}) \qquad (3.20)\\
&= \mathbb{P}\left(\|X(t_i^{(n)}) - f(t_i^{(n)})\| \leqslant r + \sum_{l=n}^{\infty} r_l, \|X(t_{i'}^{(n-1)}) - f(t_{i'}^{(n-1)})\| > r + \sum_{l=n-1}^{\infty} r_l \right)\\
&\leqslant \mathbb{P}\left(\|X(t_i^{(n)}) - f(t_i^{(n)})\| \leqslant c_2\, r, \|X(t_i^{(n)}) - X(t_{i'}^{(n-1)})\| > r_{n-1} - L\varepsilon_{n-1} \right)\\
&\leqslant \mathbb{P}\left(\|X(t_i^{(n)}) - f(t_i^{(n)})\| \leqslant c_2\, r, \|X(t_i^{(n)}) - X(t_{i'}^{(n-1)})\| > (\beta 2^{\frac{n}{2}} - L)\varepsilon_{n-1} \right),
\end{aligned}
$$

where $c_2 = 1 + \beta \sum_{l=1}^{\infty} 2^{\frac{l+1}{2}} \exp(-2^{l+1})$. We ensure $(\beta 2^{\frac{n}{2}} - L) > 0$ by choosing $\beta > L/2$. The idea is to rewrite $X(t_i^{(n)}) - X(t_{i'}^{(n-1)})$ as a sum of two terms, one expressed by $X(t_i^{(n)})$ and the other independent of $X(t_i^{(n)})$.

$\mathrm{Lip}_\rho(L)$ is invariant under orthogonal transformations. By the spectral theorem we may choose new coordinates such that the covariance matrix at $t_i^{(n)}$ is diagonal. Then the components of $X(t_i^{(n)})$ are independent. By assumption $\sigma_j(s) := \sqrt{\mathbb{E}\left[X_j(s)^2\right]} > 0$. We define the standard normal random variables

$$
Y_j(s) := \frac{X_j(s)}{\sigma_j(s)}.
$$

Note that $\mathbb{E}[Y(t_i^{(n)})Y(t_i^{(n)})^\top] = \mathrm{Id}$ holds. If $\mathbb{E}\left[(X_j(s) - X_j(t))^2\right] > 0$ we define

$$
Y_j(s,t) := \frac{X_j(s) - X_j(t)}{\sqrt{\mathbb{E}\left[(X_j(s) - X_j(t))^2\right]}}
$$

and $Y_j(s,t) := 0$ otherwise. We further define a matrix η and a random vector Z by

$$
\begin{aligned}
\eta &:= \mathbb{E}\left[Y(t_i^{(n)}, t_{i'}^{(n-1)}) Y(t_i^{(n)})^\top \right],\\
Z(t_i^{(n)}, t_{i'}^{(n-1)}) &:= Y(t_i^{(n)}, t_{i'}^{(n-1)}) - \eta Y(t_i^{(n)}).
\end{aligned}
$$

We observe that $|\eta_{jk}| \leqslant 1$ and hence in the operator norm $\|\eta\| \leqslant d$. The random vectors $Z(t_i^{(n)}, t_{i'}^{(n-1)})$ and $Y(t_i^{(n)})$ are independent because the covariance matrix is the zero matrix. By the definition of $Y(t_i^{(n)})$ we see that $Z(t_i^{(n)}, t_{i'}^{(n-1)})$ and $X(t_i^{(n)})$ are independent, too. We want to bound $\mathbb{P}(A_i^{(n)} \backslash A_{i'}^{(n-1)})$. If $t_i^{(n)} = t_{i'}^{(n-1)}$ then $\mathbb{P}(A_i^{(n)} \backslash A_{i'}^{(n-1)}) = 0$ holds. Thus we may assume that $\rho(t_i^{(n)}, t_{i'}^{(n-1)}) > 0$. (3.20) is bounded by

$$
\begin{aligned}
&\mathbb{P}\left(\|X(t_i^{(n)}) - f(t_i^{(n)})\| \leqslant c_2\, r,\ \|Y(t_i^{(n)}, t_{i'}^{(n-1)})\| > \frac{(\beta 2^{\frac{n}{2}} - L)\varepsilon_{n-1}}{c\,\rho(t_i^{(n)}, t_{i'}^{(n-1)})} \right)\\
&\leqslant \mathbb{P}\left(\|X(t_i^{(n)}) - f(t_i^{(n)})\| \leqslant c_2\, r,\ \|Z(t_i^{(n)}, t_{i'}^{(n-1)})\| + \|\eta Y(t_i^{(n)})\| > \frac{\beta 2^{n/2} - L}{c} \right)
\end{aligned}
$$

$$
\begin{aligned}
&\leqslant \mathbb{P}\left(\|X(t_i^{(n)})-f(t_i^{(n)})\|\leqslant c_2\, r,\ \|Z(t_i^{(n)},t_{i'}^{(n-1)})\|>\frac{\beta 2^{n/2}-L}{2c}\right)\\
&\quad+\mathbb{P}\left(\|X(t_i^{(n)})-f(t_i^{(n)})\|\leqslant c_2\, r,\ d\|Y(t_i^{(n)})\|>\frac{\beta 2^{n/2}-L}{2c}\right)\\
&:=I_1+I_2.
\end{aligned}
$$

Each component of $Z(t_i^{(n)},t_{i'}^{(n-1)})$ is a weighted sum of at most $d+1$ standard normal random variables with weights in $[-1,1]$. Hence the variance of each component is at most $(d+1)^2$. In the following c_l with $l\in\mathbb{N}$ will denote positive constants. By the independence of $X(t_i^{(n)})$ and $Z(t_i^{(n)},t_{i'}^{(n-1)})$ we have

$$
\begin{aligned}
I_1&=\mathbb{P}\left(\|X(t_i^{(n)})-f(t_i^{(n)})\|\leqslant c_2\, r\right)\mathbb{P}\left(\|Z(t_i^{(n)},t_{i'}^{(n-1)})\|>\frac{\beta 2^{n/2}-L}{2c}\right)\\
&\leqslant c_3\, r^d\mathbb{P}\left(\|Z(t_i^{(n)},t_{i'}^{(n-1)})\|>\frac{\beta 2^{n/2}-L}{2c}\right)\\
&\leqslant c_3\, r^d\,\frac{2d}{\sqrt{2\pi}}\frac{2\sqrt{d}(d+1)c}{\beta 2^{\frac{n}{2}}-L}\exp\left(-\frac{(\beta 2^{\frac{n}{2}}-L)^2}{8c^2 d(d+1)^2}\right)\\
&\leqslant c_4\, r^d\exp\left(-\frac{(\beta 2^{\frac{n}{2}}-L)^2}{8c^2 d(d+1)^2}\right).
\end{aligned}
$$

By the definition of $Y(t_i^{(n)})$ we have with the abbreviation $\sigma_j=\sigma_j(t_i^{(n)})$

$$
\begin{aligned}
I_2&\leqslant\int_{\{\|u-f(t_i^{(n)})\|\leqslant c_2\, r,\ \|(\frac{u_k}{\sigma_k})_k\|>\frac{\beta 2^{n/2}-L}{2dc}\}}\left(\frac{1}{2\pi}\right)^{\frac{d}{2}}\frac{1}{\sigma_1\cdots\sigma_d}e^{-\frac{1}{2}\left(\frac{u_1^2}{\sigma_1^2}+\cdots+\frac{u_d^2}{\sigma_d^2}\right)}\mathrm{d}u\\
&\leqslant\int_{\{\|u-f(t_i^{(n)})\|\leqslant c_2\, r\}}\left(\frac{1}{2\pi}\right)^{\frac{d}{2}}\frac{1}{\sigma_1\cdots\sigma_d}e^{-\frac{1}{4}\left(\frac{u_1^2}{\sigma_1^2}+\cdots+\frac{u_d^2}{\sigma_d^2}\right)}\mathrm{d}u\; e^{-\frac{1}{4}\left(\frac{(\beta 2^{\frac{n}{2}}-L)^2}{4d^2c^2}\right)}\\
&\leqslant c_5\, r^d\ \exp\left(-\frac{(\beta 2^{\frac{n}{2}}-L)^2}{16d^2c^2}\right).
\end{aligned}
$$

To sum it up

$$
\begin{aligned}
\mathbb{P}(A^{(n)})&\leqslant\mathbb{P}(A^{(n-1)})+c_6\, r^d N_n\exp\left(-\frac{(\beta 2^{\frac{n}{2}}-L)^2}{16d(d+1)^2c^2}\right)\\
&\leqslant\mathbb{P}(A^{(1)})+c_6\, r^d\sum_{k=2}^{\infty}N_k\exp\left(-\frac{(\beta 2^{\frac{k}{2}}-L)^2}{16d(d+1)^2c^2}\right)
\end{aligned}
$$

$$\leqslant c_7 r^d + c_6\, r^d \sum_{k=2}^{\infty} c_1 \,\exp\left(Q 2^{k+1} - \frac{(\beta 2^{\frac{k}{2}} - L)^2}{16 d (d+1)^2 c^2} \right).$$

We choose $\beta > \max(\widetilde{c}, L/2)$ such that $\frac{\beta^2}{16d(d+1)^2c^2} > 2Q$. Then the sum converges and the lemma follows by (3.17).

4 Asymptotic normality

This chapter presents the main results for the spectral calibration method. We proof asymptotic normality for the estimators of the volatility, the drift, the intensity and for pointwise estimators of the jump density. The joint asymptotic distribution of these estimators is derived. As it turns out the asymptotic behavior is different in the mildly ill–posed case of volatility zero and in the severely ill–posed case of positive volatility. In the mildly ill–posed case the noise grows polynomially in the frequencies while it grows exponentially in the severely ill–posed case. Through the exponential growth, the noise at the cut–off frequency is predominant in the stochastic error. Then the noise is focused in the spectral domain and unfocused in the spatial domain. For the mildly ill–posed case it is the other way around. Then the noise is more evenly distributed among the frequencies but concentrated in the spatial domain. For example, the asymptotic variance depends in the mildly ill–posed case locally on the noise of the observations, whereas for the severely ill–posed case this dependence is global. The asymptotic normality results are used to construct confidence intervals and joint confidence sets in Chapter 6. In Section 4.1, we present the asymptotic normality results. They are discussed in Section 4.2. The proofs are deferred to Section 4.3.

4.1 Main results

The starting point of the error analysis is the decomposition (2.29) into the approximation error and the stochastic error. The approximation error is deterministic and only the stochastic error can be expected to converge with appropriate scaling to a normal random variable. It is common practice to resolve this problem by undersmoothing, which means that the tuning parameter is chosen such that the approximation error becomes asymptotically negligible. For the undersmoothing the cut–off value has to grow fast enough. To obtain the exact undersmoothing conditions we first divide the order of magnitude of the stochastic error by the order of magnitude of the approximation error both in terms of ϵ_n and U_n. Then we require that the quotient tends to infinity. This leads to the condition $\epsilon_n U_n^{(2s+5)/2} \to \infty$ in the case of volatility zero and to $\epsilon_n U_n^{s+1} \exp(T\sigma^2 U_n^2/2) \to \infty$ in the case of positive volatility. Since the approximation errors are negligible by these conditions, we will focus in the following on the stochastic errors.

In the theorems, we control the supremum of $\mathcal{L}_{n,U}$ and thus the remainder term $\mathcal{R}_{n,U}$ by the conditions $\epsilon_n U_n^2 \sqrt{\log(U_n)} \to 0$ and $\epsilon_n U_n^2 \exp(T\sigma^2 U_n^2/2) \to 0$ for $\sigma = 0$ and $\sigma > 0$, respectively. Then the asymptotic distribution of the stochastic errors $\int_0^1 \Delta\psi_n(Uu)w(u)\mathrm{d}u$ is governed by the linearized stochastic errors $\int_0^1 \mathcal{L}_{n,U}(u)w(u)\mathrm{d}u$ and the remainder term $\int_0^1 \mathcal{R}_{n,U}(u)w(u)\mathrm{d}u$ is asymptotically negligible. In the case $\sigma = 0$ the stronger condition $\epsilon_n U_n^{5/2} \to 0$ is assumed, which is needed for the stochastic errors to converge to zero.

In the results on asymptotic normality we will also include the estimator $\hat{\mu}(0)$ of the jump density at zero. This only makes sense by our smoothness assumption on μ since there is no way of detecting jumps of height zero. Unlike for points $x \neq 0$ it will turn out that not the weight function w_μ^1 determines the asymptotic distribution but the effective weight function

$$w_0(u) := w_\mu^1(u) + w_\sigma^1(u)\int_{-1}^1 v^2 w_\mu^1(v)\mathrm{d}v/2 - w_\lambda^1(u)\int_{-1}^1 w_\mu^1(v)\mathrm{d}v.$$

The first theorem states the joint asymptotic normality result for the mildly ill–posed case of volatility zero.

Theorem 4.1. *Let $\sigma = 0$. Let δ be continuous at $T\gamma, x_1 + T\gamma, \ldots, x_m + T\gamma$ and let $\mathcal{F}\delta^2 \in L^1(\mathbb{R})$. For $j = 1, \ldots, m$ let $x_j \in \mathbb{R}\setminus\{0\}$ be distinct and let $V_0, W_0, W_{x_1} \ldots, W_{x_m}$ be independent Brownian motions. If $\epsilon_n U_n^{5/2} \to 0$ and $\epsilon_n U_n^{(2s+5)/2} \to \infty$ as $n \to \infty$, then*

$$\frac{1}{\epsilon_n}\begin{pmatrix} U_n^{+1/2} & \Delta\hat{\sigma}^2 \\ U_n^{-1/2} & \Delta\hat{\gamma} \\ U_n^{-3/2} & \Delta\hat{\lambda} \\ U_n^{-5/2} & \Delta\hat{\mu}(0) \\ U_n^{-5/2} & \Delta\hat{\mu}(x_1) \\ & \vdots \\ U_n^{-5/2} & \Delta\hat{\mu}(x_m) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} d(0) & \int_0^1 u^2 w_\sigma^1(u)\mathrm{d}W_0(u) \\ d(0) & \int_0^1 u^2 w_\gamma^1(u)\mathrm{d}V_0(u) \\ d(0) & \int_0^1 u^2 w_\lambda^1(u)\mathrm{d}W_0(u) \\ d(0) & \int_0^1 u^2 w_0(u)\mathrm{d}W_0(u)/(2\pi) \\ d(x_1) & \int_0^1 u^2 w_\mu^1(u)\mathrm{d}W_{x_1}(u)/(2\pi) \\ & \vdots \\ d(x_n) & \int_0^1 u^2 w_\mu^1(u)\mathrm{d}W_{x_m}(u)/(2\pi) \end{pmatrix},$$

as $n \to \infty$, where $d(x) := 2\sqrt{\pi}\delta(x + T\gamma)\exp(T(\lambda - \gamma))/T$.

Remark 4.2. The theorem is formulated in terms of the exponentially weighted jump density $\mu(x) = e^x\nu(x)$. By the continuous mapping theorem results on μ can be reformulated in terms of ν by multiplying with e^{-x_j} in the respective lines.

Proof. We write $\Delta\hat{\gamma}$, $\Delta\hat{\lambda}$ and $\Delta\hat{\mu}(x)$ similarly as in (2.29) for $\Delta\hat{\sigma}^2$:

$$\Delta\hat{\gamma} = -\Delta\hat{\sigma}^2 + \frac{2}{U}\int_0^1 \mathrm{Im}(\mathcal{F}\mu(Uu))w_\gamma^1(u)\mathrm{d}u + \frac{2}{U}\int_0^1 \mathrm{Im}(\Delta\psi_n(Uu))w_\gamma^1(u)\mathrm{d}u, \tag{4.1}$$

$$\Delta\hat{\lambda} = \frac{\Delta\hat{\sigma}^2}{2} + \Delta\hat{\gamma} - 2\int_0^1 \mathrm{Re}(\mathcal{F}\mu(Uu))w_\lambda^1(u)\mathrm{d}u - 2\int_0^1 \mathrm{Re}(\Delta\psi_n(Uu))w_\lambda^1(u)\mathrm{d}u, \tag{4.2}$$

$$\begin{aligned}\Delta\hat{\mu}(x) = {} & U\mathcal{F}^{-1}\left[\Delta\psi_n(Uu)w_\mu^1(u)\right](Ux) \\ & + \frac{\Delta\hat{\sigma}^2}{2}U\mathcal{F}^{-1}\left[(Uu - i)^2 w_\mu^1(u)\right](Ux) - i\Delta\hat{\gamma}U\mathcal{F}^{-1}\left[(Uu - i)w_\mu^1(u)\right](Ux) \\ & + \Delta\hat{\lambda}U\mathcal{F}^{-1}\left[w_\mu^1(u)\right](Ux) - U\mathcal{F}^{-1}\left[(1 - w_\mu^1(u))\mathcal{F}\mu(Uu)\right](Ux). \end{aligned} \tag{4.3}$$

In (4.1) we can substitute $\Delta\hat{\sigma}^2$ using (2.29) and obtain two error terms involving $\mathcal{F}\mu$ and two error terms involving $\Delta\psi_n$. By similar substitutions in (4.2) and (4.3) we see that all error terms either involve $\mathcal{F}\mu$ or $\Delta\psi_n$, which we will call approximation errors and stochastic errors, respectively.

The undersmoothing $\epsilon_n U_n^{(2s+5)/2} \to \infty$ is equivalent to $U_n^{-(s+3)} = o(\epsilon_n U_n^{-1/2})$. The approximation error of $\hat{\sigma}^2$ decays by (4.29) below as $U_n^{-(s+3)}$ and thus is asymptotically negligible. The three approximation errors $2U^{-1}\int_0^1 \mathrm{Im}(\mathcal{F}\mu(Uu))w_\gamma^1(u)\mathrm{d}u$ of $\hat{\gamma}$, $2\int_0^1 \mathrm{Re}(\mathcal{F}\mu(Uu))w_\lambda^1(u)\mathrm{d}u$ of $\hat{\lambda}$ and $U\mathcal{F}^{-1}\left[(1-w_\mu^1(u))\mathcal{F}\mu(Uu)\right](Ux)$ of $\hat{\mu}(x)$ can be bounded similarly as done in (4.30), (4.31) and (4.32) below and are asymptotically negligible, too. Since $\hat{\sigma}^2$ converges with a faster rate than $\hat{\gamma}$ and $\hat{\gamma}$ converges with a faster rate than $\hat{\lambda}$, the errors $\Delta\hat{\sigma}^2$ in (4.1) and in (4.2) as well as $\Delta\hat{\gamma}$ in (4.2) are asymptotically negligible. For $x \neq 0$ we can apply the Riemann–Lebesgue lemma to the second, the third and the fourth error term in (4.3) and we see that they are of order $o_{\mathbb{P}}(\epsilon_n U_n^{5/2})$. For $x = 0$ due to the symmetry of w_μ^1 the third term vanishes asymptotically but the second and the fourth term do not. The error terms of $\hat{\mu}(x)$ we have to consider are in the case $x \neq 0$

$$U\mathcal{F}^{-1}\left[\Delta\psi_n(Uu)w_\mu^1(u)\right](Ux) = \frac{U}{2\pi}2\int_0^1 w_\mu^1(u)\mathrm{Re}\left(\Delta\psi_n(Uu)e^{-iUux}\right)\mathrm{d}u$$

and in the case $x = 0$

$$\begin{aligned}&\mathcal{F}^{-1}\left[\Delta\psi_n(Uu))w_\mu^1(u)\right](0) + \int_0^1 \mathrm{Re}(\Delta\psi_n(Uu))w_\sigma^1(u)\mathrm{d}u\,\mathcal{F}^{-1}\left[u^2 w_\sigma^1(u)\right](0)\\ &- 2\int_0^1 \mathrm{Re}(\Delta\psi_n(Uu))w_\lambda^1(u)\mathrm{d}u\,\mathcal{F}^{-1}\left[w_\mu^1(u)\right](0)\\ = &\frac{1}{2\pi}2\int_0^1 \mathrm{Re}(\Delta\psi_n(Uu))w_0(u)\mathrm{d}u.\end{aligned}$$

By assumption (2.17) on the order of the weight functions, w_σ^1, w_γ^1, w_λ^1 and w_μ^1 are continuous and bounded, especially they are Riemann–integrable and in $L^\infty([-1,1])$. As the main technical step, Lemma 4.5 shows the convergence of the linearized stochastic errors. The remainder terms are asymptotically negligible by Lemma 4.11. □

Next we consider the case $\sigma > 0$. Let δ be in $L^\infty(\mathbb{R})$ and $\|\delta\|_{L^2(\mathbb{R})} > 0$. We set

$$d := \sqrt{2}\|\delta\|_{L^2(\mathbb{R})}\exp(T(\lambda - \gamma - \sigma^2/2))T^{-2}\sigma^{-2} \tag{4.4}$$

and define by $W_{n,U} + iV_{n,U} := 2d^{-1}\int_0^1 \mathcal{L}_{n,U}(u)\mathrm{d}u$ the real–valued random variables $W_{n,U}$ and $V_{n,U}$. By Lemma 4.8 below

$$\frac{1}{\epsilon_n \exp(T\sigma^2 U^2/2)}\begin{pmatrix} W_{n,U}\\ V_{n,U}\end{pmatrix} \xrightarrow{d} \begin{pmatrix} W\\ V\end{pmatrix} \tag{4.5}$$

as $U \to \infty$, where W and V are independent standard normal random variables.

The following theorem treats the stochastic errors in the case of positive volatility. Since the theorem contains no statement on the approximation errors, the condition (2.17) on the order of the weight functions may be omitted.

Theorem 4.3. *Let $\sigma > 0$ and $\delta \in L^\infty(\mathbb{R})$. Assume for the cut–off value $U_n \to \infty$ and $\epsilon_n U_n^2 \exp(T\sigma^2 U_n^2/2) \to 0$ as $n \to \infty$. Let $w_\sigma^1, w_\gamma^1, w_\lambda^1, w_\mu^1 : [0,1] \to \mathbb{R}$ be Riemann–integrable, in $L^\infty([0,1])$ and continuous at one. Then for $x \in \mathbb{R}$*

$$\frac{1}{\epsilon_n \exp(T\sigma^2 U_n^2/2)} \begin{pmatrix} 2\int_0^1 \operatorname{Re}(\Delta\psi_n(U_n u))w_\sigma^1(u)\mathrm{d}u - d\,w_\sigma^1(1)W_{n,U_n} \\ 2\int_0^1 \operatorname{Im}(\Delta\psi_n(U_n u))w_\gamma^1(u)\mathrm{d}u - d\,w_\gamma^1(1)\,V_{n,U_n} \\ 2\int_0^1 \operatorname{Re}(\Delta\psi_n(U_n u))w_\lambda^1(u)\mathrm{d}u - d\,w_\lambda^1(1)W_{n,U_n} \\ \mathcal{F}^{-1}\left[\Delta\psi_n(U_n u)w_\mu^1(u)\right](U_n x) - d\,w_\mu^1(1)\,Z_{n,U_n}(x)/(2\pi) \end{pmatrix} \xrightarrow{\mathbb{P}} 0,$$

as $n \to \infty$, where $Z_{n,U}(x) := \cos(Ux)W_{n,U} + \sin(Ux)V_{n,U}$.

Proof. The main technical step is provided by Lemma 4.9, which treats the convergence of the linearized stochastic errors. The remainder terms are asymptotically negligible by Lemma 4.10. To see the first line we set $x_1 = x_2 = 0$, $w_1 \equiv 1$ and $w_2 = w_\sigma^1$ in Lemma 4.9 and $w_U = w_\sigma^1$ in Lemma 4.10. The second and third line follow analogously. In order to derive the last line we observe

$$\mathcal{F}^{-1}\left[\Delta\psi_n(Uu)w_\mu^1(u)\right](Ux) = 2\int_0^1 \operatorname{Re}(\Delta\psi_n(Uu)e^{-iUux})w_\mu^1(u)\mathrm{d}u/(2\pi)$$

and apply Lemma 4.9 with $x_1 = 0$, $x_2 = x$, $w_1 \equiv 1$ and $w_2 = w_\mu^1$, . The remainder term vanishes by setting $w_U(u) = w_\mu^1(u)e^{-iUux}$ in Lemma 4.10. □

The assumption $\mathcal{T} \in \mathcal{G}_s(R, \sigma_{\max})$ restricts σ to the interval $[0, \sigma_{\max}]$. The condition $\epsilon_n U_n^2 \exp(T\sigma^2 U_n^2/2) \to 0$ is especially fulfilled if $U_n \leqslant \bar{\sigma}^{-1}(2\log(\epsilon_n^{-1})/T)^{1/2}$ for any $\bar{\sigma} > \sigma_{\max}$. For the estimation it suffices to know some upper bound $\sigma_{\max}$ of σ. The theorem shows that regardless whether one undersmoothes or not the stochastic errors converge with appropriate scaling to normal random variables. For the statement on asymptotic normality we have to undersmooth.

In many situations the volatility σ is known or can be estimated easily. The volatility is preserved under a change to an equivalent measure so that it is the same for the risk–neutral and the real–world measure even if the real–world price process is only assumed to be a semimartingale. Then one of the methods for volatility estimation from high frequency data in the presence of jumps can be used to estimate the volatility. Cont and Tankov (2004b) also need to fix the volatility for their calibration method of exponential Lévy models in advance since their method chooses only among measures of Lévy processes equivalent to a prior measure. They suggest using historical data or an earlier calibrated model for the choice of the prior and thus also of the volatility. In the following we will assume either that the volatility σ is known or that we have a sufficiently good estimator of the volatility. The volatility is needed for choosing the growth rate of the cut–off value such that on the one hand we undersmooth and on the other hand the remainder term is asymptotically negligible. In practice there are other ways to determine the cut–off value, where knowledge about σ is not needed. So while σ plays a special role in the theory, it can be treated as the other parameters in the simulations and in the empirical study in Chapter 7. But for our theoretical results we fix a growth rate of the cut–off value U_n. To control the remainder term we choose U_n such that $\epsilon_n U_n^2 \exp(T\sigma^2 U_n^2/2) \to 0$ as $n \to \infty$. We also assume the undersmoothing condition $\epsilon_n U_n^{s+1} \exp(T\sigma^2 U_n^2/2) \to \infty$ as $n \to \infty$. A smoothness parameter $s \geqslant 2$ is implicitly assumed so that both condition can be satisfied simultaneously. A possible choice of U_n is

$$U_n := \left(\frac{2}{T\sigma^2}\log\left(\frac{\epsilon_n^{-1}}{\log(\epsilon_n^{-1})^\alpha}\right)\right)^{1/2}, \tag{4.6}$$

where $\alpha \in (1, (s+1)/2)$. Then it holds

$$\epsilon_n U_n^\beta \exp(T\sigma^2 U_n^2/2) \to \begin{cases} \infty & \beta > 2\alpha \\ \left(\frac{2}{T\sigma^2}\right)^\alpha & \beta = 2\alpha \\ 0 & \beta < 2\alpha \end{cases}$$

as $n \to \infty$. Especially the term diverges for $\beta = s+1$ and converges to zero for $\beta = 2$ so that both conditions on U_n are fulfilled.

Next we state the joint asymptotic normality result for the severely ill–posed case of positive volatility.

Theorem 4.4. *Let $\sigma > 0$ and $\delta \in L^\infty(\mathbb{R})$. Let the cut–off value U_n be chosen such that $\epsilon_n U_n^2 \exp(T\sigma^2 U_n^2/2) \to 0$ and $\epsilon_n U_n^{s+1} \exp(T\sigma^2 U_n^2/2) \to \infty$ as $n \to \infty$. Then*

$$\frac{1}{\epsilon_n \exp(T\sigma^2 U_n^2/2)} \begin{pmatrix} U_n^2 & \Delta\hat{\sigma}^2 & - & d\, w_\sigma^1(1) W_{n,U_n} \\ U_n & \Delta\hat{\gamma} & - & d\, w_\gamma^1(1) V_{n,U_n} \\ & \Delta\hat{\lambda} & - & d\, w_\lambda^1(1) W_{n,U_n} \\ U_n^{-1} & \Delta\hat{\mu}(0) & - & d\, w_0(1) W_{n,U_n}/(2\pi) \\ U_n^{-1} & \Delta\hat{\mu}(x) & - & d\, w_\mu^1(1) Z_{n,U_n}(x)/(2\pi) \end{pmatrix} \xrightarrow{\mathbb{P}} 0,$$

as $n \to \infty$, where $x \in \mathbb{R}\setminus\{0\}$, $Z_{n,U}(x) := \cos(Ux)W_{n,U} + \sin(Ux)V_{n,U}$ and d is given by (4.4)*.*

Proof. The undersmoothing condition $\epsilon_n U_n^{s+1} \exp(T\sigma^2 U_n^2/2) \to \infty$ yields $U_n^{-(s+3)} = o(\epsilon_n U_n^{-2} \exp(T\sigma^2 U_n^2/2))$ so that the approximation error of $\hat{\sigma}^2$ vanishes. A similar reasoning applies to the approximation errors of the other estimators. Since $\hat{\sigma}^2$ converges with a faster rate than $\hat{\gamma}$ and $\hat{\gamma}$ with a faster rate than $\hat{\lambda}$ the leading stochastic error terms are given in Theorem 4.3 and the convergence of the first three lines follows by this theorem. For $x \neq 0$ all stochastic errors in (4.3) are negligible except the first one. We obtain the convergence in the last line by Theorem 4.3. We observe that $\mathcal{F}^{-1}[uw_\mu^1(u)](0) = 0$, since w_μ^1 is symmetric. For $x=0$ the relevant stochastic error terms are

$$\begin{aligned}
&\mathcal{F}^{-1}\left[\Delta\psi_n(Uu)w_\mu^1(u)\right](0)+\int_0^1 \mathrm{Re}(\Delta\psi_n(Uu))w_\sigma^1(u)\mathrm{d}u\,\mathcal{F}^{-1}\left[u^2 w_\mu^1(u)\right](0)\\
&-2\int_0^1 \mathrm{Re}(\Delta\psi_n(Uu))w_\lambda^1(u)\mathrm{d}u\,\mathcal{F}^{-1}\left[w_\mu^1(u)\right](0)\\
=&\frac{1}{2\pi}2\int_0^1 \mathrm{Re}(\Delta\psi_n(Uu))w_0(u)\mathrm{d}u.
\end{aligned}$$

We apply Lemma 4.9 with $x_1=x_2=0$, $w_1\equiv 1$ and $w_2=w_0$ to this term. The remainder term is asymptotically negligible by Lemma 4.10. This shows the convergence in the next to last line. □

4.2 Discussion of the results

Theorems 4.1 and 4.4 include the asymptotic distribution of $\hat\sigma^2$, which may be used for testing the hypotheses $H_0:\sigma=\sigma_0$, see Section 6.2. If σ is known, we can set $\hat\sigma^2=\sigma^2$. Then the statements of the theorems hold with w_σ^1 constant to zero. The estimation method can give negative values for $\hat\sigma^2$, $\hat\lambda$ and $\hat\nu(x)$. By a postprocessing step the estimated values can be corrected to be non–negative.

In Theorem 4.1 the noise level δ enters only locally into the asymptotic variance, whereas in Theorems 4.3 and 4.4 the asymptotic variance depends on the L^2–norm of δ through the factor d. In fact, for $\sigma=0$ it is possible to estimate γ and λ directly from local properties of the option function $\mathcal{O}$ at γT as remarked in Belomestny and Reiß (2006a). This local dependence on the noise level resembles some similarity to deconvolution, for instance, to the case of ordinary smooth error density Fan (1991a) or to the case of symmetric stable error densities whose characteristic function decreases slower than the characteristic function of the Cauchy distribution van Es and Uh (2004). In both cases the density of the observations enters locally into the asymptotic variance. For the weight functions the local and global dependence is vice versa. In Theorem 4.1 the weight functions w_σ^1, w_γ^1, w_λ^1, w_0^1 and w_μ^1 enter globally into the asymptotic variance while in Theorems 4.3 and 4.4 only the values of the weight functions at their endpoints appear in the asymptotic variance.

The asymptotic variance depends on the time to maturity. For positive volatility this dependence is through d in (4.4). The martingale condition is equivalent to $\sigma^2/2+\gamma-\lambda+\int_{-\infty}^{\infty}e^x\nu(x)\mathrm{d}x=0$, especially it holds that $\lambda-\gamma-\sigma^2/2\geqslant 0$ with equality if and only if $\lambda=0$, that is in the Black–Scholes case. In the case of positive volatility the asymptotic variance grows quadratic as $T\to 0$ and, if the jump intensity λ is positive, exponentially as $T\to\infty$. In view of the Lévy–Khintchine representation $T\sigma^2$, $T\gamma$, $T\lambda$ and $T\mu$ are the natural objects to estimate. This explains the factor T^{-1} in the asymptotic variance in Theorem 4.1. In Theorems 4.3 and 4.4 there is an additional factor of $T^{-1}\sigma^{-2}$ entering through d defined in (4.4). This factor stems from the fact that the mass assigned to the highest frequencies close to cut–off value is proportional to $T^{-1}\sigma^{-2}$.

For $w_\sigma^1(1)$, $w_\gamma^1(1)$, $w_\lambda^1(1)$, $w_\mu^1(1) \in \mathbb{R}\setminus\{0\}$, Theorem 4.3 describes the asymptotic distribution of the leading stochastic error terms of $\hat{\sigma}^2$, $\hat{\gamma}$, $\hat{\lambda}$ and $\hat{\mu}(x)$, $x \neq 0$, i.e., the other stochastic error terms are of smaller order. Theorem 4.4 describes the asymptotic distribution of the estimation errors. Both theorems are for the case of positive volatility, where the noise in the frequency domain is exponentially heteroscedastic, so that the highest frequency, that is the cut–off frequency U, dominates the stochastic error. Then this cut–off frequency U can be seen in the asymptotic distribution of $\Delta\hat{\mu}$ through the oscillating process $Z_{n,U}$. The variances in Theorems 4.3 and 4.4 converge by (4.5) and by the definition of $Z_{n,U}(x)$. If one only considers the stochastic errors of $\hat{\sigma}^2$, $\hat{\gamma}$, $\hat{\lambda}$ and $\hat{\mu}(0)$, then the covariances converge, too. But for $x \neq 0$ the covariance of the stochastic errors of $\hat{\mu}(x)$ and of $\hat{\sigma}^2$ does not converge. The same holds for the covariance of the stochastic errors of $\hat{\mu}(x)$ and $\hat{\gamma}$ as well as $\hat{\mu}(x)$ and $\hat{\lambda}$. The phenomenon that the covariances do not convergence comes from the fact that the stochastic error centers more and more at the cut–off frequency. The sequence of cut–off values has a crucial influence on the covariance. For estimators of the generalized distribution function of the jump density this is likely to lead to a similar dependence on the sequence of cut–off values as observed in van Es and Uh (2005) for deconvolution with supersmooth errors.

4.3 Proof of the asymptotic normality

4.3.1 The linearized stochastic errors

The linearized stochastic errors are of the form $\int_0^1 f_j(u)\mathcal{L}_{n,U}(u)\mathrm{d}u$, where f_j with $j = 1,\dots,n$ are Riemann–integrable functions in $L^\infty([0,1])$. Next we will show that these are jointly normal distributed. Almost surely $\mathcal{L}_{n,U}$ is continuous. Thus, almost surely the $f_j\mathcal{L}_{n,U}$ are Riemann–integrable and almost surely

$$\frac{1}{m}\sum_{k=1}^{m} f_j(k/m)\mathcal{L}_{n,U}(k/m) \to \int_0^1 f_j(u)\mathcal{L}_{n,U}(u)\mathrm{d}u$$

as $m \to \infty$. Let $C > 0$ be such that $\|f_j\|_\infty \leqslant C$ for all $j = 1,\dots,n$. For each m the n sums are joint, centered normal random variables. For $m \to \infty$ the covariance matrix converges by the dominated convergence theorem with the dominating function $C^2 \sup_{u\in[0,1]} |\mathcal{L}_{n,U}(u)|^2$, where $\sup_{u\in[0,1]} |\mathcal{L}_{n,U}(u)|^2$ is an integrable random variable by Proposition 3.3. Thus, the characteristic function converges pointwise. By Lévy's continuity theorem this shows that the sums convergence jointly in distribution to normal random variables. So $\int_0^1 f_j(u)\mathcal{L}_{n,U}(u)\mathrm{d}u$ are jointly normal distributed.

For a fixed cut–off value U the linearized stochastic errors are jointly normal distributed. So the natural question is whether the appropriately scaled covariance matrix converges for $U \to \infty$.

Let $w_j, w_k : [0,1] \to \mathbb{R}$ be Riemann–integrable functions in $L^\infty([0,1])$. It holds

$$
\begin{aligned}
&T\int_0^1 w_j(u)\mathcal{L}_{n,U}(u)e^{-iUux_j}\mathrm{d}u\\
&= \epsilon_n U^2 e^{-T(\sigma^2/2+\gamma-\lambda)}\int_0^1 f_U(u)\int_{-\infty}^{\infty} e^{iUu(x-\theta_j)+T\sigma^2U^2u^2/2}\delta(x)\mathrm{d}W(x)\mathrm{d}u,
\end{aligned}
$$

where

$$
f_U(u) := \frac{w_j(u)(-u^2+iu/U)}{\exp(T\mathcal{F}\mu(Uu))} \tag{4.7}
$$

and $\theta_j := x_j + T\sigma^2 + T\gamma$. We define analogously

$$
g_U(u) := \frac{w_k(u)(-u^2+iu/U)}{\exp(T\mathcal{F}\mu(Uu))} \tag{4.8}
$$

and $\theta_k := x_k + T\sigma^2 + T\gamma$. We extend f_U and g_U by zero outside the interval $[0,1]$.

Since $\mathbb{E}\left[\sup_{u\in[-1,1]}|\mathcal{L}_{n,U}(u)|^2\right] < \infty$, we may apply Fubini's theorem and then we apply the Itô isometry to obtain

$$
\begin{aligned}
&T^2e^{2T(\sigma^2/2+\gamma-\lambda)}\mathbb{E}\left[\int_0^1 w_j(u)\mathcal{L}_{n,U}(u)e^{-iUux_j}\mathrm{d}u\overline{\int_0^1 w_k(v)\mathcal{L}_{n,U}(v)e^{-iUvx_k}\mathrm{d}v}\right]\\
&= \epsilon_n^2U^4\int_0^1\int_0^1\int_{-\infty}^{\infty} f_U(u)e^{iUu(x-\theta_j)+T\sigma^2U^2u^2/2}\\
&\quad\overline{g_U(v)e^{iUv(x-\theta_k)+T\sigma^2U^2v^2/2}}\delta(x)^2\mathrm{d}x\mathrm{d}u\mathrm{d}v.
\end{aligned} \tag{4.9}
$$

To separate real and imaginary part we will also need

$$
\begin{aligned}
&T^2e^{2T(\sigma^2/2+\gamma-\lambda)}\mathbb{E}\left[\int_0^1 w_j(u)\mathcal{L}_{n,U}(u)e^{-iUux_j}\mathrm{d}u\int_0^1 w_k(v)\mathcal{L}_{n,U}(v)e^{-iUvx_k}\mathrm{d}v\right]\\
&= \epsilon_n^2U^4\int_0^1\int_0^1\int_{-\infty}^{\infty} f_U(u)e^{iUu(x-\theta_j)+T\sigma^2U^2u^2/2}\\
&\quad g_U(v)e^{iUv(x-\theta_k)+T\sigma^2U^2v^2/2}\delta(x)^2\mathrm{d}x\mathrm{d}u\mathrm{d}v.
\end{aligned} \tag{4.10}
$$

Lemma 4.5. *Let $\sigma = 0$. For $j = 1,\dots,m$ let $x_j \in \mathbb{R}$ and let $w_j : [0,1] \to \mathbb{R}$ be Riemann–integrable functions in $L^\infty([0,1])$. Let δ be continuous at $x_1+T\gamma$, $x_2+T\gamma,\dots,x_m+T\gamma$ and let $\mathcal{F}\delta^2 \in L^1(\mathbb{R})$. Let $W_{x_1},\dots,W_{x_m},V_{x_1},\dots,V_{x_m}$ be Brownian motions. If $x_j = x_k$ let $W_{x_j} = W_{x_k}$ and $V_{x_j} = V_{x_k}$ otherwise let the Brownian motions be distinct. Let the set $\{W_{x_1},\dots,W_{x_m},V_{x_1},\dots,V_{x_m}\}$ consist of independent Brownian motions. Then*

$$
\frac{1}{\epsilon_nU^{3/2}}\int_0^1 w_j(u)\mathcal{L}_{n,U}(u)e^{-iUux_j}\mathrm{d}u
$$

converge jointly in distribution to

$$\frac{\sqrt{\pi}\delta(x_j + T\gamma)}{T\exp(T(\gamma - \lambda))}\left(\int_0^1 u^2 w_j(u)\mathrm{d}W_{x_j}(u) + i\int_0^1 u^2 w_j(u)\mathrm{d}V_{x_j}(u)\right)$$

as $U \to \infty$.

Proof. We will first consider the case $x_j = x_k$. We have seen that

$$T^2 e^{2T(\gamma-\lambda)}\mathbb{E}\left[\int_0^1 w_j(u)\mathcal{L}_{n,U}(u)e^{-iUux_j}\mathrm{d}u\overline{\int_0^1 w_k(v)\mathcal{L}_{n,U}(v)e^{-iUvx_j}\mathrm{d}v}\right]$$
$$= \epsilon_n^2 U^4 \int_{-\infty}^{\infty} \mathcal{F}f_U(U(x-\theta_j))\overline{\mathcal{F}g_U(U(x-\theta_j))}\delta(x)^2\mathrm{d}x,$$

where f_U and g_U are defined as in (4.7) and (4.8), respectively, and $\theta_j = x_j + T\gamma$,

$$= \epsilon_n^2 U^3 \int_{-\infty}^{\infty} \mathcal{F}f_U(y)\overline{\mathcal{F}g_U(y)\delta(y/U+\theta_j)^2}\mathrm{d}y.$$

We notice that $\mathcal{F}\delta^2 \in L^1(\mathbb{R})$ implies $\delta^2 \in L^\infty(\mathbb{R})$ and we obtain by the Plancherel identity

$$= 2\pi\epsilon_n^2 U^3 \int_0^1 f_U(u)\overline{(g_U(v) * \mathcal{F}^{-1}(\delta(y/U+\theta_j)^2)(v))(u)}\mathrm{d}u, \tag{4.11}$$

since the support of f_U is $[0,1]$. Because we are only interested in the limit $U \to \infty$, we may assume $U \geqslant 1$. By the Riemann–Lebesgue lemma $\mathcal{F}\mu(u)$ tends to zero as $|u| \to \infty$. The factor $f_U(u)$ converges for each $u \in [0,1]$ to $-u^2 w_j(u)$ as $U \to \infty$ and the functions are dominated by a constant independent of U. In order to apply dominated convergence it suffices that the second factor is dominated by a constant independent of U and converges stochastically with respect to the Lebesgue measure on $\mathbb{R}$.

$$(g_U(v) * \mathcal{F}^{-1}(\delta(y/U+\theta_j)^2)(v))(u) = \int_{-\infty}^{\infty} g_U(u-v)\mathcal{F}^{-1}(\delta(y+\theta_j)^2)(Uv)U\mathrm{d}v$$

By assumption $\mathcal{F}\delta^2$ lies in $L^1(\mathbb{R})$ and so does $\mathcal{F}^{-1}(\delta(y+\theta_j)^2)$. A dominating constant is $\sqrt{2}\|w_k\|_\infty \exp(T\|\mathcal{F}\mu\|_\infty)\,\|\mathcal{F}^{-1}(\delta(y+\theta_j)^2)\|_{L^1(\mathbb{R})}$. It holds

$$\int_{-\infty}^{\infty} \mathcal{F}^{-1}(\delta(y+\theta_j)^2)(Uv)U\mathrm{d}v = \int_{-\infty}^{\infty} \mathcal{F}^{-1}(\delta(y+\theta_j)^2)(v)\mathrm{d}v$$
$$= \mathcal{F}\mathcal{F}^{-1}(\delta(y+\theta_j)^2)(0) = \delta(\theta_j)^2. \tag{4.12}$$

$\delta_U(v) := \mathcal{F}^{-1}(\delta(y+\theta_j)^2)(Uv)U$ is the multiple of what is called approximate identity or nascent delta function (Grafakos, 2004, Definition 1.2.15). For $h \in L^1(\mathbb{R})$ and an approximate identity δ_n we have that $h * \delta_n$ converges to h in $L^1(\mathbb{R})$ as $n \to \infty$. Thus, $(-v^2 w_k(v)\mathbf{1}_{[0,1]}(v)) * \delta_U(v)(u)$ converges to $-u^2 w_k(u)\mathbf{1}_{[0,1]}(u)\delta(\theta_j)^2$ for $U \to \infty$ in $L^1(\mathbb{R})$

(Grafakos, 2004, p. 28) and in particular stochastically (Klenke, 2007, Definition 6.2). If $u \neq 0$, then there is a neighborhood of u where $g_U(u) + u^2 w_k(u)\mathbf{1}_{[0,1]}(u)$ converges uniformly to zero. $(g_U(v)+v^2 w_k(v)\mathbf{1}_{[0,1]}(v))*\delta_U(v)(u)$ converges almost surely and in particular stochastically to zero. Thus, $g_U(v)*\delta_U(v)(u)$ converges to $-u^2 w_k(u)\mathbf{1}_{[0,1]}(u)\delta(\theta_j)^2$ stochastically with respect to the Lebesgue measure on $\mathbb{R}$.

We obtain under the limit $U \to \infty$ by the dominated convergence theorem

$$\begin{aligned}
&\lim_{U\to\infty} \frac{1}{\epsilon_n^2 U^3}\mathbb{E}\left[\int_0^1 w_j(u)\mathcal{L}_{n,U}(u)e^{-iUux_j}\mathrm{d}u\overline{\int_0^1 w_k(v)\mathcal{L}_{n,U}(v)e^{-iUvx_j}\mathrm{d}v}\right]\\
&= \frac{2\pi\delta(\theta_j)^2}{T^2\exp(2T(\gamma-\lambda))}\int_{-\infty}^{\infty}\left(-u^2 w_j(u)\mathbf{1}_{[0,1]}(u)\right)\left(-u^2 w_k(u)\mathbf{1}_{[0,1]}(u)\right)\mathrm{d}u\\
&= \frac{2\pi\delta(\theta_j)^2}{T^2\exp(2T(\gamma-\lambda))}\int_0^1 u^4 w_j(u)w_k(u)\mathrm{d}u. \qquad (4.13)
\end{aligned}$$

Without taking the complex conjugate in (4.11) we obtain

$$\begin{aligned}
&T^2 e^{2T(\gamma-\lambda)}\mathbb{E}\left[\int_0^1 w_j(u)\mathcal{L}_{n,U}(u)e^{-iUux_j}\mathrm{d}u\int_0^1 w_k(v)\mathcal{L}_{n,U}(v)e^{-iUvx_j}\mathrm{d}v\right]\\
&= 2\pi\int_{-\infty}^{\infty} f_U(u)\overline{(\overline{g_U(-v)} * \mathcal{F}^{-1}(\delta(y/U+\theta_j)^2)(v))(u)}\mathrm{d}u.
\end{aligned}$$

The same argumentation as before leads to

$$\begin{aligned}
&\lim_{U\to\infty}\mathbb{E}\left[\int_0^1 w_j(u)\mathcal{L}_{n,U}(u)e^{-iUux_j}\mathrm{d}u\int_0^1 w_k(v)\mathcal{L}_{n,U}(v)e^{-iUvx_j}\mathrm{d}v\right]\\
&= \frac{2\pi\delta(\theta_j)^2}{T^2\exp(2T(\gamma-\lambda))}\int_{-\infty}^{\infty}\left(-u^2 w_j(u)\mathbf{1}_{[0,1]}(u)\right)\left(-u^2 w_k(-u)\mathbf{1}_{[0,1]}(-u)\right)\mathrm{d}u\\
&= 0. \qquad (4.14)
\end{aligned}$$

We combine (4.13) and (4.14) to obtain

$$\begin{aligned}
&\lim_{U\to\infty}\frac{1}{\epsilon_n^2 U^3}\mathbb{E}\left[\operatorname{Re}\int_0^1 \frac{w_j(u)\mathcal{L}_{n,U}(u)}{\exp(iUux_j)}\mathrm{d}u\operatorname{Re}\int_0^1\frac{w_k(u)\mathcal{L}_{n,U}(u)}{\exp(iUux_j)}\mathrm{d}v\right]\\
&= \lim_{U\to\infty}\frac{1}{\epsilon_n^2 U^3}\mathbb{E}\left[\operatorname{Im}\int_0^1 \frac{w_j(u)\mathcal{L}_{n,U}(u)}{\exp(iUux_j)}\mathrm{d}u\operatorname{Im}\int_0^1\frac{w_k(u)\mathcal{L}_{n,U}(u)}{\exp(iUux_j)}\mathrm{d}v\right]\\
&= \frac{\pi\delta(\theta_j)^2}{T^2\exp(2T(\gamma-\lambda))}\int_0^1 u^4 w_j(u)w_k(u)\mathrm{d}u.
\end{aligned}$$

From (4.13) and (4.14) it also follows that the covariance between real and imaginary part vanishes asymptotically.

In the case $x_j \neq x_k$, we have to show that the covariance vanishes asymptotically. Without loss of generality we assume $x_j < x_k$. We define $\theta := (\theta_j + \theta_k)/2$, yielding

$\theta_j < \theta < \theta_k$. We have

$$\begin{aligned}
&\frac{T^2 e^{2T(\gamma-\lambda)}}{\epsilon_n^2 U^3}\mathbb{E}\left[\int_0^1 w_j(u)\mathcal{L}_{n,U}(u)e^{-iUux_j}\mathrm{d}u\overline{\int_0^1 w_k(v)\mathcal{L}_{n,U}(v)e^{-iUvx_k}\mathrm{d}v}\right]\\
&= \int_{-\infty}^{\infty}\mathcal{F}f_U(U(x-\theta_j))\overline{\mathcal{F}g_U(U(x-\theta_k))}\delta(x)^2 U\mathrm{d}x\\
&= \int_{-\infty}^{\infty}\mathcal{F}f_U(y+U(\theta-\theta_j))\overline{\mathcal{F}g_U(y+U(\theta-\theta_k))}\delta(y/U+\theta)^2\mathrm{d}y.
\end{aligned}$$

By the Plancherel identity and by the dominated convergence theorem

$$\mathcal{F}f_U \to \mathcal{F}\left(-u^2 w_j(u)\right) \quad \text{and} \quad \mathcal{F}g_U \to \mathcal{F}\left(-u^2 w_k(u)\right)$$

in $L^2(\mathbb{R})$ for $U \to \infty$ and especially the $L^2(\mathbb{R})$ norms converge. From the assumption $\mathcal{F}\delta^2 \in L^1(\mathbb{R})$ follows that $\delta^2 \in L^\infty(\mathbb{R})$. By the Cauchy–Schwarz inequality

$$\begin{aligned}
&\lim_{U\to\infty}\left|\int_{-\infty}^0 \mathcal{F}f_U(y+U(\theta-\theta_j))\overline{\mathcal{F}g_U(y+U(\theta-\theta_k))}\delta(y/U+\theta)^2\mathrm{d}y\right|\\
&\leqslant \lim_{U\to\infty}\|\delta\|_\infty^2\,\|\mathcal{F}f_U\|_{L^2(\mathbb{R})}\left(\int_{-\infty}^{U(\theta-\theta_k)}|\mathcal{F}g_U(y)|^2\,\mathrm{d}y\right)^{1/2} = 0.
\end{aligned}$$

A similar calculation shows that the integral over $(0,\infty)$ converges to zero and consequently

$$\lim_{U\to\infty}\frac{T^2 e^{2T(\gamma-\lambda)}}{\epsilon_n^2 U^3}\mathbb{E}\left[\int_0^1 \frac{w_j(u)\mathcal{L}_{n,U}(u)}{\exp(iUux_j)}\mathrm{d}u\overline{\int_0^1 \frac{w_k(v)\mathcal{L}_{n,U}(v)}{\exp(iUvx_k)}\mathrm{d}v}\right] = 0.$$

The same way follows

$$\lim_{U\to\infty}\frac{T^2 e^{2T(\gamma-\lambda)}}{\epsilon_n^2 U^3}\mathbb{E}\left[\int_0^1 \frac{w_j(u)\mathcal{L}_{n,U}(u)}{\exp(iUux_j)}\mathrm{d}u\int_0^1 \frac{w_k(v)\mathcal{L}_{n,U}(v)}{\exp(iUvx_k)}\mathrm{d}v\right] = 0.$$

The $1/(\epsilon_n U^{3/2})\int_0^1 w_j(u)\mathcal{L}_{n,U}(u)e^{-iUux_j}\mathrm{d}u$ are centered normal random variables and their covariance matrix converges to the covariance matrix of the claimed limit. Thus, the characteristic function converges pointwise. By Lévy's continuity theorem this shows the convergence in distribution. $\square$

Lemma 4.6. *Let $\sigma > 0$ and $\delta \in L^\infty(\mathbb{R})$. Let $w_U, \widetilde{w}_U \in L^\infty([0,1],\mathbb{C})$ be Riemann–integrable and let there be a constant $C > 0$ such that for all $U \geqslant 1$ $\|w_U\|_\infty, \|\widetilde{w}_U\|_\infty \leqslant C$. Let there be $a, \widetilde{a} : [1,\infty) \to \mathbb{C}$ such that the condition*

$$\lim_{\eta\to 0}\sup_{U\geqslant 1}\sup_{u\in[1-\eta/U,1]}|w_U(u)-a(U)| = 0 \tag{4.15}$$

and the corresponding condition for $\widetilde{w}_U$ and $\widetilde{a}$ hold. Then

$$\lim_{U\to\infty} \frac{1}{\epsilon_n^2 \exp(T\sigma^2 U^2)} \mathbb{E}\left[\int_0^1 w_U(u)\mathcal{L}_{n,U}(u)\mathrm{d}u \int_0^1 \widetilde{w}_U(v)\mathcal{L}_{n,U}(v)\mathrm{d}v\right] = 0,$$

$$\lim_{U\to\infty}\Bigg(\frac{1}{\epsilon_n^2 \exp(T\sigma^2 U^2)} \mathbb{E}\left[\int_0^1 w_U(u)\mathcal{L}_{n,U}(u)\mathrm{d}u \overline{\int_0^1 \widetilde{w}_U(v)\mathcal{L}_{n,U}(v)\mathrm{d}v}\right] \\ - \frac{a(U)\overline{\widetilde{a}(U)}\int_{-\infty}^{\infty}\delta(y)^2\mathrm{d}y}{\exp(2T(\sigma^2/2+\gamma-\lambda))T^4\sigma^4}\Bigg) = 0.$$

Remark 4.7. Obviously $a(U) := w_U(1)$ is the only possible definition. Thus, a describes the dependence of w_U on U at one.

Proof. We notice that (4.9) applies to the complex–valued functions and yields for $w_j := w_U$ and $w_k := \widetilde{w}_U$ with the definitions (4.7) and (4.8) of f_U and g_U, respectively, and with $\theta := T\sigma^2 + T\gamma$

$$\begin{aligned} &T^2 e^{2T(\sigma^2/2+\gamma-\lambda)}\mathbb{E}\left[\int_0^1 w_U(u)\mathcal{L}_{n,U}(u)\mathrm{d}u\overline{\int_0^1 \widetilde{w}_U(v)\mathcal{L}_{n,U}(v)\mathrm{d}v}\right] \\ &= \epsilon_n^2 U^4 \int_{-\infty}^{\infty} \mathcal{F}(f_U(u)e^{T\sigma^2U^2u^2/2})(U(x-\theta))\overline{\mathcal{F}(g_U(v)e^{T\sigma^2U^2v^2/2})(U(x-\theta))}\delta(x)^2\mathrm{d}x, \\ &= 2\pi\epsilon_n^2 U^3 \int_{-\infty}^{\infty} f_U(u)e^{T\sigma^2U^2u^2/2}\left(\overline{g_U(v)}e^{T\sigma^2U^2v^2/2} * \overline{\mathcal{F}^{-1}(\delta(y/U+\theta)^2)(v)}\right)(u)\mathrm{d}u \\ &= 2\pi\epsilon_n^2 U^4 \int_0^1\int_0^1 f_U(u)\overline{g_U(v)\mathcal{F}^{-1}(\delta(y+\theta)^2)(U(u-v))}e^{T\sigma^2U^2(u^2+v^2)/2}\mathrm{d}u\mathrm{d}v. \end{aligned} \tag{4.16}$$

For all $\eta > 0$ we have

$$\begin{aligned} &\lim_{U\to\infty} T\sigma^2U^2e^{-T\sigma^2U^2/2}\int_{1-\eta/U}^1 ue^{T\sigma^2U^2u^2/2}\mathrm{d}u \\ &= \lim_{U\to\infty} e^{-T\sigma^2U^2/2}\left[e^{T\sigma^2U^2u^2/2}\right]_{1-\eta/U}^1 = 1 - \lim_{U\to\infty} e^{-T\sigma^2U\eta+T\sigma^2\eta^2/2} = 1. \end{aligned} \tag{4.17}$$

For the product of two such sequences we obtain for all $\eta > 0$

$$\lim_{U\to\infty} T^2\sigma^4U^4e^{-T\sigma^2U^2}\int_{1-\eta/U}^1\int_{1-\eta/U}^1 uve^{T\sigma^2U^2(u^2+v^2)/2}\mathrm{d}u\mathrm{d}v = 1. \tag{4.18}$$

Likewise

$$\lim_{U\to\infty} T^2\sigma^4U^4e^{-T\sigma^2U^2}\int_0^1\int_0^1 uve^{T\sigma^2U^2(u^2+v^2)/2}\mathrm{d}u\mathrm{d}v = 1 \tag{4.19}$$

holds. We scale the integral in (4.16) appropriately:

$$\lim_{U\to\infty}\Big(T^2\sigma^4U^4e^{-T\sigma^2U^2}\int_0^1\int_0^1 f_U(u)\overline{g_U(v)} \tag{4.20}$$

$$\overline{\mathcal{F}^{-1}(\delta(y+\theta)^2)(U(u-v))}e^{T\sigma^2U^2(u^2+v^2)/2}\mathrm{d}u\mathrm{d}v-a(U)\overline{\widetilde{a}(U)}\frac{1}{2\pi}\int_{-\infty}^{\infty}\delta(y)^2\mathrm{d}y\Big)$$

$$=\lim_{U\to\infty}\Big(T^2\sigma^4U^4e^{-T\sigma^2U^2}\int_0^1\int_0^1 uve^{T\sigma^2U^2(u^2+v^2)/2} \tag{4.21}$$

$$\overline{\mathcal{F}^{-1}(\delta(y+\theta)^2)(U(u-v))}f_U(u)\overline{g_U(v)}/(uv)\mathrm{d}u\mathrm{d}v-a(U)\overline{\widetilde{a}(U)}\frac{1}{2\pi}\int_{-\infty}^{\infty}\delta(y)^2\mathrm{d}y\Big)$$

We recall that in the Gaussian white noise model we assumed δ to be in $L^2(\mathbb{R})$. Since $\overline{\mathcal{F}^{-1}(\delta(y+\theta)^2)(U(u-v))}f_U(u)\overline{g_U(v)}/(uv)$ is bounded in $L^\infty([0,1]^2)$ independently of U for $U\geqslant 1$ and since the difference between (4.19) and (4.18) is zero, only the integral over $[1-\eta/U,1]^2$ contributes to the limit. For all $\eta>0$ the limit equals

$$=\lim_{U\to\infty}\Big(T^2\sigma^4U^4e^{-T\sigma^2U^2}\int_{1-\eta/U}^1\int_{1-\eta/U}^1 uve^{T\sigma^2U^2(u^2+v^2)/2} \tag{4.22}$$

$$\overline{\mathcal{F}^{-1}(\delta(y+\theta)^2)(U(u-v))}f_U(u)\overline{g_U(v)}/(uv)\mathrm{d}v\mathrm{d}u-a(U)\overline{\widetilde{a}(U)}\frac{1}{2\pi}\int_{-\infty}^{\infty}\delta(y)^2\mathrm{d}y\Big)$$

$$=0,$$

which can be seen the following way. $\mathcal{F}^{-1}(\delta(y+\theta)^2)$ is continuous and $|U(u-v)|\leqslant\eta$ for all $u,v\in[1-\eta/U,1]$. So by choosing η small enough $\mathcal{F}^{-1}(\delta(y+\theta)^2)(U(u-v))$ gets arbitrarily close to $\mathcal{F}^{-1}(\delta(y+\theta)^2)(0)=(1/2\pi)\int_{-\infty}^{\infty}\delta(y)^2\mathrm{d}y$. By (4.15), $w_U(u)$ tends to $a(U)$ and $\widetilde{w}_U(v)$ tends to $\widetilde{a}(U)$ for η tending to zero. By choosing U, large the factor $(-u+i/U)/\exp(T\mathcal{F}\mu(Uu))$ gets close to minus one for all $u\in[1-\eta/U,1]$. Thus, for small η and large U the term $f_U(u)\overline{g_U(v)}/(uv)$ is close to $a(U)\overline{\widetilde{a}(U)}$ for all $u,v\in[1-\eta/U,1]$.

Rescaling (4.16) and taking the limit $U\to\infty$ leads to

$$\lim_{U\to\infty}\Big(\frac{1}{\epsilon_n^2\exp(T\sigma^2U^2)}\mathbb{E}\left[\int_0^1 w_U(u)\mathcal{L}_{n,U}(u)\mathrm{d}u\overline{\int_0^1\widetilde{w}_U(v)\mathcal{L}_{n,U}(v)\mathrm{d}v}\right]$$

$$-\frac{a(U)\overline{\widetilde{a}(U)}\int_{-\infty}^{\infty}\delta(y)^2\mathrm{d}y}{\exp(2T(\sigma^2/2+\gamma-\lambda))T^4\sigma^4}\Big)$$

$$=\lim_{U\to\infty}\Big(\frac{2\pi U^4\exp(-T\sigma^2U^2)}{T^2\exp(2T(\sigma^2/2+\gamma-\lambda))}\int_0^1\int_0^1 f_U(u)\overline{g_U(v)}$$

$$\overline{\mathcal{F}^{-1}(\delta(y+\theta_0)^2)(U(u-v))}e^{T\sigma^2U^2(u^2+v^2)/2}\mathrm{d}u\mathrm{d}v$$

$$-\frac{2\pi}{\exp(2T(\sigma^2/2+\gamma-\lambda))}\frac{a(U)\overline{\widetilde{a}(U)}}{T^4\sigma^4 2\pi}\int_{-\infty}^{\infty}\delta(y)^2\mathrm{d}y\Big)=0, \tag{4.23}$$

where we used that (4.20) is zero. By (4.10) with $y = U(x-\theta)$ we have

$$
\begin{aligned}
&T^2 e^{2T(\sigma^2/2+\gamma-\lambda)}\mathbb{E}\left[\int_0^1 w_U(u)\mathcal{L}_{n,U}(u)\mathrm{d}u \int_0^1 \widetilde{w}_U(v)\mathcal{L}_{n,U}(v)\mathrm{d}v\right] \\
&= \epsilon_n^2 U^3 \int_{-\infty}^{\infty} \mathcal{F}(f_U(u)e^{T\sigma^2U^2u^2/2})(y)\overline{\mathcal{F}(\overline{g_U(-v)}e^{T\sigma^2U^2v^2/2})(y)}\delta(y/U+\theta)^2\mathrm{d}y \\
&= 2\pi\epsilon_n^2 U^3 \int_{-\infty}^{\infty} f_U(u)e^{T\sigma^2U^2u^2/2}\left(g_U(-v)e^{T\sigma^2U^2v^2/2} * \overline{\mathcal{F}^{-1}(\delta(y/U+\theta)^2)(v)}\right)(u)\mathrm{d}u \\
&= 2\pi\epsilon_n^2 U^4 \int_0^1\int_{-1}^0 f_U(u)g_U(-v)\overline{\mathcal{F}^{-1}(\delta(y+\theta)^2)(U(u-v))}e^{T\sigma^2U^2(u^2+v^2)/2}\mathrm{d}v\mathrm{d}u \\
&= 2\pi\epsilon_n^2 U^4 \int_0^1\int_0^1 uve^{T\sigma^2U^2(u^2+v^2)/2}\mathcal{F}^{-1}(\delta(y+\theta)^2)(-U(u+v))\frac{f_U(u)g_U(v)}{uv}\mathrm{d}v\mathrm{d}u.
\end{aligned}
$$

Rescaling as in (4.23) leads to

$$
\begin{aligned}
&\lim_{U\to\infty}\frac{1}{\epsilon_n^2\exp(T\sigma^2U^2)}\mathbb{E}\left[\int_0^1 w_U(u)\mathcal{L}_{n,U}(u)\mathrm{d}u\int_0^1\widetilde{w}_U(v)\mathcal{L}_{n,U}(v)\mathrm{d}v\right] \\
=&\lim_{U\to\infty}\frac{2\pi U^4\exp(-T\sigma^2U^2)}{T^2\exp(2T(\sigma^2/2+\gamma-\lambda))}\int_0^1\int_0^1 uve^{T\sigma^2U^2(u^2+v^2)/2} \\
&\mathcal{F}^{-1}(\delta(y+\theta)^2)(-U(u+v))f_U(u)g_U(v)/(uv)\mathrm{d}u\mathrm{d}v = 0, \qquad (4.24)
\end{aligned}
$$

since $\mathcal{F}^{-1}(\delta(y+\theta)^2)(u)\to 0$ for $|u|\to\infty$. □

Lemma 4.8. *Let $\sigma > 0$ and $\delta \in L^\infty(\mathbb{R})$. Let $x_0 \in \mathbb{R}$ and for $j = 1,\ldots,n$ let $w_j : [0,1] \to \mathbb{R}$ be continuous at one, Riemann–integrable and in $L^\infty([0,1])$. Then*

$$\frac{1}{\epsilon_n\exp(T\sigma^2U^2/2)}\int_0^1 w_j(u)\mathcal{L}_{n,U}(u)e^{-iUux_0}\mathrm{d}u$$

converge jointly in distribution to

$$\frac{\|\delta\|_{L^2(\mathbb{R})}w_j(1)}{\sqrt{2}\exp(T(\sigma^2/2+\gamma-\lambda))T^2\sigma^2}(W+iV)$$

as $U\to\infty$, where W and V are independent standard normal random variables.

Proof. This is a consequence of Lemma 4.6. We define $w_U(u) := w_j(u)/\exp(iUux_0)$ and $\widetilde{w}_U(u) := w_k(u)/\exp(iUux_0)$. Further we set $a(U) := w_j(1)/\exp(iUx_0)$ and $\widetilde{a}(U) := w_k(1)/\exp(iUx_0)$. Then we apply Lemma 4.6. Condition (4.15) is satisfied since w_j and w_k are continuous at one and since $\exp(-iUux_0) = \exp(-iUx_0)\exp(iU(1-u)x_0)$, where $U(1-u) \leqslant \eta$ for $u \in [1-\eta/U, 1]$. We note that $a(U)\overline{\widetilde{a}(U)} = w_j(1)w_k(1)$ is real. By Lemma 4.6 the covariances converge to the covariances of the claimed limit. The convergence in distribution follows by Lévy's continuity theorem. □

Lemma 4.9. *Let $\sigma > 0$ and $\delta \in L^\infty(\mathbb{R})$. Let $w_1, w_2 : [0,1] \to \mathbb{R}$ be Riemann–integrable, in $L^\infty([0,1])$ and continuous at one. Let $x_1, x_2 \in \mathbb{R}$ and denote $x_2 - x_1$ by φ. Then*

$$\frac{1}{\epsilon_n e^{T\sigma^2 U^2/2}}\left(w_1(1)\int_0^1 \frac{w_2(u)\mathcal{L}_{n,U}(u)}{e^{iUux_2}}\mathrm{d}u - \frac{w_2(1)}{e^{iU\varphi}}\int_0^1 \frac{w_1(u)\mathcal{L}_{n,U}(u)}{e^{iUux_1}}\mathrm{d}u\right) \xrightarrow{\mathbb{P}} 0,$$

as $U \to \infty$.

Proof. The proof relies on Lemma 4.6. We define

$$w_U(u) := \frac{w_1(1)w_2(u)}{\exp(iUux_2)} - \frac{w_2(1)w_1(u)}{\exp(iU\varphi)\exp(iUux_1)}.$$

$w_U(u)$ fulfills condition (4.15) with $a(U) = 0$ for all $U \geqslant 1$. Lemma 4.6 yields

$$\lim_{U\to\infty} \frac{1}{\epsilon_n^2 \exp(T\sigma^2U^2)}\mathbb{E}\left[\left|\int_0^1 w_U(u)\mathcal{L}_{n,U}(u)\mathrm{d}u\right|^2\right] = 0$$

and the statement follows by Lévy's continuity theorem. □

4.3.2 The remainder term

In this section, we show that the contribution of the remainder term to the estimation vanishes asymptotically. We recall that the remainder term $\mathcal{R}_{n,U}$ depends on the characteristic triplet.

Lemma 4.10. *Let $\sigma_0 > 0$. Let $w_U \in L^\infty([0,1],\mathbb{C})$ be Riemann–integrable and let there be a constant $C > 0$ such that $\|w_U\|_\infty \leqslant C$ for all $U \geqslant 1$. If $\epsilon_n U_n^2 \exp(T\sigma_0^2 U_n^2/2) \to 0$ as $n \to \infty$, then for all characteristic triplets with $\sigma \leqslant \sigma_0$*

$$\frac{1}{\epsilon_n \exp(T\sigma_0^2 U_n^2/2)}\int_0^1 w_{U_n}(u)\mathcal{R}_{n,U_n}(u)\mathrm{d}u \xrightarrow{\mathbb{P}} 0, \qquad \textit{as } n \to \infty.$$

Proof. By the identity $\mathcal{R}_{n,U}(u) = (1/T)\log_{\geqslant\kappa^U(u)}(1+T\mathcal{L}_{n,U}(u)) - \mathcal{L}_{n,U}(u)$ with $\kappa^U(u) \leqslant 1/2$, we have to show that with the abbreviation $U = U_n$

$$\frac{1}{\epsilon_n \exp(T\sigma_0^2U^2/2)}\int_0^1 w_U(u)\Big(\log_{\geqslant\kappa^U(u)}(1+T\mathcal{L}_{n,U}(u)) - T\mathcal{L}_{n,U}(u)\Big)\mathrm{d}u \tag{4.25}$$

converges in probability to zero.

For $z \in \mathbb{C}$ holds $\log(1+z) - z = O(|z|^2)$ as $|z| \to 0$. We define g by $g(z) := (\log(1+z)-z)/|z|^2$ for $z \neq 0$ and $g(0) := 0$. There are M and $\eta > 0$ such that $|g(z)| \leqslant M$ for all $|z| \leqslant \eta$. We may assume that $\eta \leqslant 1/2$. If the logarithm in the definition of g is replaced by the trimmed logarithm $\log_{\geqslant\kappa}$ with $\kappa \in (0,1/2]$ then g remains unchanged for $|z| \leqslant 1/2$. Thus, the statement holds uniformly for all $g_\kappa(z) := (\log_{\geqslant\kappa}(1+z)-z)/|z|^2$ with $\kappa \in (0,1/2]$.

By Proposition 3.3 we have $\sup_{u\in[-1,1]} |\mathcal{L}_{n,U}(u)| \xrightarrow{\mathbb{P}} 0$. Let $\tau > 0$ be given. Eventually we have

$$\mathbb{P}\left(\exists u \in [-1,1] : |\log_{\geqslant\kappa^U(u)}(1+T\mathcal{L}_{n,U}(u)) - T\mathcal{L}_{n,U}(u)| > MT^2|\mathcal{L}_{n,U}(u)|^2\right)$$
$$\leqslant \mathbb{P}\left(T \sup_{u\in[-1,1]} |\mathcal{L}_{n,U}(u)| > \eta\right) < \tau.$$

Except on a set with probability less than τ we have eventually

$$\frac{1}{\epsilon_n \exp(T\sigma_0^2 U^2/2)} \left|\int_0^1 w_U(u)\left(\log_{\geqslant\kappa^U(u)}(1+T\mathcal{L}_{n,U}(u)) - T\mathcal{L}_{n,U}(u)\right)\mathrm{d}u\right|$$
$$\leqslant \frac{MT^2}{\epsilon_n \exp(T\sigma_0^2 U^2/2)} \int_0^1 |w_U(u)\mathcal{L}_{n,U}(u)^2|\mathrm{d}u. \tag{4.26}$$

Hence (4.25) converges in probability to zero if (4.26) converges in probability to zero. The convergence

$$\frac{1}{\epsilon_n \exp(T\sigma_0^2 U^2/2)} \int_0^1 |w_U(u)\mathcal{L}_{n,U}(u)^2|\mathrm{d}u \to 0$$

holds even in L^1 since

$$\frac{1}{\epsilon_n \exp(T\sigma_0^2 U^2/2)}\mathbb{E}\left[\int_0^1 |w_U(u)\mathcal{L}_{n,U}(u)^2|\mathrm{d}u\right] \tag{4.27}$$
$$\leqslant \frac{C}{\epsilon_n \exp(T\sigma_0^2 U^2/2)}\mathbb{E}\left[\int_0^1 |\mathcal{L}_{n,U}(u)^2|\mathrm{d}u\right]$$
$$\leqslant \frac{C}{\epsilon_n \exp(T\sigma_0^2 U^2/2)}$$
$$\int_0^1 \left|\frac{\epsilon_n\, iUu(1+iUu)}{T(1+iUu(1+iUu)\mathcal{F}\mathcal{O}(Uu))}\right|^2 \mathbb{E}\left[\left|\int_{-\infty}^{\infty} e^{iUux}\delta(x)\mathrm{d}W(x)\right|^2\right]\mathrm{d}u$$
$$\leqslant \frac{C}{\epsilon_n \exp(T\sigma_0^2 U^2/2)} \int_0^1 \frac{\epsilon_n^2(U^2+U^4)u\exp(T\sigma^2U^2u^2)\|\delta\|^2_{L^2(\mathbb{R})}}{T^2\exp(2T(\sigma^2/2+\gamma-\lambda)-2T\|\mathcal{F}\mu\|_\infty)}\mathrm{d}u, \tag{4.28}$$

for $\sigma = 0$ this converges to zero and for $\sigma > 0$ we further calculate,

$$= \frac{C\epsilon_n(1+U^2)\|\delta\|^2_{L^2(\mathbb{R})}\int_0^1 2T\sigma^2U^2u\exp(T\sigma^2U^2u^2)\mathrm{d}u}{\exp(T\sigma_0^2U^2/2)2T^3\sigma^2\exp(2T(\sigma^2/2+\gamma-\lambda)-2T\|\mathcal{F}\mu\|_\infty)}$$
$$= \frac{C\epsilon_n(1+U^2)\|\delta\|^2_{L^2(\mathbb{R})}(\exp(T\sigma^2U^2)-1)}{\exp(T\sigma_0^2U^2/2)2T^3\sigma^2\exp(2T(\sigma^2/2+\gamma-\lambda)-2T\|\mathcal{F}\mu\|_\infty)}$$
$$\leqslant \frac{C\|\delta\|^2_{L^2(\mathbb{R})}\epsilon_n(1+U^2)(\exp(T\sigma_0^2U^2/2))}{2T^3\sigma^2\exp(2T(\sigma^2/2+\gamma-\lambda)-2T\|\mathcal{F}\mu\|_\infty)} \to 0$$

as $n\to\infty$. Thus, (4.25) converges in probability to zero. □

Lemma 4.11. *Let $w_U \in L^\infty([0,1],\mathbb{C})$ be Riemann–integrable and let there be a constant $C > 0$ such that $\|w_U\|_\infty \leqslant C$ for all $U \geqslant 1$. If $U_n \to \infty$ and $\epsilon_n U_n^{5/2} \to 0$ as $n \to \infty$, then for all characteristic triplets with $\sigma = 0$*

$$\frac{1}{\epsilon_n U_n^{3/2}} \int_0^1 w_{U_n}(u)\mathcal{R}_{\epsilon_n,U_n}(u)\mathrm{d}u \xrightarrow{\mathbb{P}} 0,$$

as $n \to \infty$.

Proof. We follow the proof of Lemma 4.10. We have $\sup_{u\in[-1,1]} |\mathcal{L}_{n,U}(u)| \xrightarrow{\mathbb{P}} 0$ by Proposition 3.3. We set $\sigma_0 = 0$ and divide by $U^{3/2}$ in (4.25) and (4.26). Then we use that (4.27) is bounded by (4.28), where we set $\sigma_0 = \sigma = 0$ and divide by $U^{3/2}$ again. We obtain

$$\frac{1}{\epsilon_n U^{3/2}}\mathbb{E}\left[\int_0^1 |w_U(u)\mathcal{L}_{n,U}(u)^2|\mathrm{d}u\right] \leqslant \frac{C\epsilon_n\,(U^{1/2}+U^{5/2})\|\delta\|^2_{L^2(\mathbb{R})}}{T^2\exp(T(2(\gamma-\lambda)-2\|\mathcal{F}\mu\|_\infty))} \to 0$$

as $\epsilon_n \to 0$, which implies the desired convergence. □

4.3.3 The approximation errors

The approximation error can be controlled as in Belomestny and Reiß (2006a) using the order conditions (2.17) on the weight functions. The characteristic triplet $\mathcal{T} = (\sigma^2,\gamma,\mu)$ was assumed to be contained in $\mathcal{G}_s(R,\sigma_{\max})$, especially μ is s–times weakly differentiable and we have $\max_{0\leqslant k\leqslant s} \|\mu^{(k)}\|_{L^2(\mathbb{R})} \leqslant R$, $\|\mu^{(s)}\|_\infty \leqslant R$.

We use $(iu)^s\mathcal{F}\mu(u) = \mathcal{F}\mu^{(s)}(u)$ and the Plancherel identity to bound the approximation error by

$$\begin{aligned}
&\left|\frac{2}{U^2}\int_0^1 \mathrm{Re}(\mathcal{F}\mu(Uu))w^1_\sigma(u)\mathrm{d}u\right| = \frac{1}{U^2}\left|\int_{-1}^1 \mathcal{F}\mu(Uu)w^1_\sigma(u)\mathrm{d}u\right| \\
&= \frac{2\pi}{U^2}\left|\int_{-\infty}^{\infty} \mu^{(s)}(x/U)U^{-1}\overline{\mathcal{F}^{-1}(w^1_\sigma(u)/(iUu)^s)(x)}\mathrm{d}x\right| \\
&\leqslant U^{-(s+3)}\|\mu^{(s)}\|_\infty\|\mathcal{F}(w^1_\sigma(u)/u^s)\|_{L^1(\mathbb{R})}. \qquad (4.29)
\end{aligned}$$

Analogously we obtain

$$\left|\frac{2}{U}\int_0^1 \mathrm{Im}(\mathcal{F}\mu(Uu))w^1_\gamma(u)\mathrm{d}u\right| \leqslant U^{-(s+2)}\|\mu^{(s)}\|_\infty\|\mathcal{F}(w^1_\gamma(u)/u^s)\|_{L^1(\mathbb{R})}, \qquad (4.30)$$

$$\left|2\int_0^1 \mathrm{Re}(\mathcal{F}\mu(Uu))w^1_\lambda(u)\mathrm{d}u\right| \leqslant U^{-(s+1)}\|\mu^{(s)}\|_\infty\|\mathcal{F}(w^1_\lambda(u)/u^s)\|_{L^1(\mathbb{R})}. \qquad (4.31)$$

The last error term in (4.3) can be bounded by

$$
\begin{aligned}
&\left|U\mathcal{F}^{-1}\left[(1-w_\mu^1(u))\mathcal{F}\mu(Uu)\right](Ux)\right| = \frac{U}{2\pi}\left|\int_{-\infty}^{\infty}(1-w_\mu^1(u))\mathcal{F}\mu(Uu)e^{-iUux}\mathrm{d}u\right| \\
&= \frac{1}{2\pi U^{s-1}}\left|\int_{-\infty}^{\infty}\overline{\frac{1-w_\mu^1(u)}{u^s}e^{iUux}}\mathcal{F}\mu^{(s)}(Uu)\mathrm{d}u\right| \\
&= U^{-s}\left|\int_{-\infty}^{\infty}\overline{\mathcal{F}^{-1}\left(\frac{1-w_\mu^1(u)}{u^s}e^{iUux}\right)(y)}\mu^{(s)}\left(\frac{y}{U}\right)\mathrm{d}y\right| \\
&\leqslant \frac{\|\mu^{(s)}\|_\infty}{2\pi U^s}\left\|\mathcal{F}\left(\frac{1-w_\mu^1(u)}{u^s}\right)\right\|_{L^1(\mathbb{R})}. \qquad (4.32)
\end{aligned}
$$

5 Uniformity with respect to the underlying probability measure

In order to obtain honest confidence sets, we have to consider asymptotic normality not only for one single probability measure but rather over a class of probability measures. To this end, we will extend the asymptotic normality results of Chapter 4. The results hold for each characteristic triplet $\mathcal{T} \in \mathcal{G}_s(R, \sigma_{\max})$ specified in Definition 2.4. The speed of convergence might depend on $\mathcal{T}$. To make statements on confidence sets and on hypotheses tests it is useful to control the speed of convergence uniformly over a class of characteristic triplets. We fix some arbitrarily slowly decreasing function h with $h(u) \to 0$ as $|u| \to \infty$. We will show uniform convergence for the class $\mathcal{H}_s(R, \sigma_{\max})$ consisting of all characteristic triplets in $\mathcal{G}_s(R, \sigma_{\max})$ satisfying the additional conditions

$$\|\mathcal{F}\mu\|_\infty \leqslant R \quad \text{and} \quad |\mathcal{F}\mu(u)| \leqslant R\,h(u), \quad \forall u \in \mathbb{R}\,. \tag{5.1}$$

The first condition can easily be ensured by $\|\mu\|_{L^1(\mathbb{R})} \leqslant R$. For $h(u) = |u|^{-1}$ the second condition can be ensured by $\|\mu'\|_{L^1(\mathbb{R})} \leqslant R$. For each characteristic triplet in $\mathcal{G}_s(R, \sigma_{\max})$ the function $\mathcal{F}\mu(u)$ tends to zero as $|u| \to 0$ by the Riemann–Lebesgue lemma. Especially for each $\mathcal{T} \in \mathcal{G}_s(R, \sigma_{\max})$ there are h and $\widetilde{R} > 0$ such that $\mathcal{T} \in \mathcal{H}_s(\widetilde{R}, \sigma_{\max})$.

In the case $\sigma > 0$ some covariances do not converge. We show uniform convergence for the joint distribution only in such cases, where the covariances do converge. As it turns out it is also important that the covariance matrix of the limit is nondegenerate. We cover uniform convergence of $\hat{\sigma}^2$, $\hat{\gamma}$, $\hat{\lambda}$, $\hat{\mu}(x)$ and of the pair $(\hat{\gamma}, \hat{\lambda})$.

In Section 5.1, we describe our general approach to obtain uniform convergence with respect to the underlying probability measure. Section 5.2 and Section 5.3 cover the specific aspects for volatility zero and for positive volatility, respectively. Section 5.4 treats uniformity in the probability measure for the remainder term.

5.1 General approach

Before we state the uniform version of our asymptotic normality results, we begin with some definitions. A sequence of random variables X_n converges to X *in total variation* if

$$\sup_B |\mathbb{P}(X_n \in B) - \mathbb{P}(X \in B)| \to 0$$

as $n \to \infty$, where the supremum is taken over all measurable sets B. For sequences $X_{\vartheta,n}$, $\vartheta \in \Theta$, we say that they converge to X *in total variation uniformly over* Θ if

$$\sup_{\vartheta\in\Theta} \sup_{B} |\mathbb{P}(X_{\vartheta,n} \in B) - \mathbb{P}(X \in B)| \to 0$$

as $n \to \infty$. The extension of convergence in total variation to convergence in total variation uniformly over Θ is canonic. How to define convergence in distribution uniformly over Θ is less clear. By the Portmanteau theorem there are a number of equivalent definitions of convergence in distribution. One definition is that a sequence of random variables X_n converges to X *in distribution* if for all Borel sets B with $\mathbb{P}(X \in \partial B) = 0$ we have

$$|\mathbb{P}(X_n \in B) - \mathbb{P}(X \in B)| \to 0 \qquad \text{as } n \to \infty.$$

This definition is particularly suitable for applications to confidence sets. We have such applications in mind and thus define convergence in distribution uniform over Θ based on this definition. We say for $X_{\vartheta,n}$, $\vartheta \in \Theta$, and X with values in some metric space that $X_{\vartheta,n}$ converge to X *in distribution uniformly over* Θ if for all Borel sets B with $\mathbb{P}(X \in \partial B) = 0$ we have

$$\sup_{\vartheta\in\Theta} |\mathbb{P}(X_{\vartheta,n} \in B) - \mathbb{P}(X \in B)| \to 0 \qquad \text{as } n \to \infty.$$

This definition is directly applicable for the construction of honest confidence sets, but does not seem to be established in the literature. A different approach is taken by van der Vaart and Wellner (1996) who consider Donsker theorems which are uniform in the underlying probability measure. They base their notion of uniformity on the characterization of convergence in distribution through bounded, Lipschitz continuous functions, see the beginning of their Section 2.8.2. The next theorem is the uniform version of our asymptotic normality results.

Theorem 5.1. *Let the assumptions of Theorem 4.1 be fulfilled and let δ be continuous and positive. Each marginal convergence in Theorem 4.1 is a uniform convergence in distribution over $\mathcal{H}_s(R, 0)$ if the standard deviation is positive and both sides are divided by it.*

Let $\sigma > 0$ and let the assumptions of Theorem 4.4 be fulfilled. Each marginal convergence in Theorem 4.4 is a uniform convergence in distribution over all $\mathcal{T} \in \mathcal{H}_s(R, \sigma_{\max})$ with volatility σ if the standard deviation is positive and both sides are divided by it.

Remark 5.2. In both cases uniform convergence in distribution does also hold for $\hat{\gamma}$ and $\hat{\lambda}$ jointly. In the standard deviation on the left side γ and λ may be replaced by their estimators.

Remark 5.3. Asymptotic equivalence can be defined through the total variation norm uniformly over Θ, see Theorem 2 in Le Cam and Yang (2000). Consequently, the concept of uniform convergence either in total variation or in distribution is compatible with the asymptotic equivalence of the regression and the Gaussian white noise model.

The following lemma may be seen as a generalization of Slutsky's lemma for uniform convergence. It is the key step to show the uniform convergence in distribution in Theorem 5.1.

Lemma 5.4. *Let $X_{\vartheta,n}$, $Y_{\vartheta,n}$, $\vartheta \in \Theta$, $n \in \mathbb{N}$, and X be random vectors such that $X_{\vartheta,n}$ converge to X in total variation uniformly over Θ and $\sup_{\vartheta\in\Theta} \mathbb{P}(|Y_{\vartheta,n}| \geqslant \delta) \to 0$ as $n \to \infty$ for all $\delta > 0$. Let $Z_{\vartheta,n}$ be random variables with $\sup_{\vartheta\in\Theta} \mathbb{P}(|Z_{\vartheta,n} - 1| \geqslant \delta) \to 0$ for all $\delta > 0$. Then $Z_{\vartheta,n}X_{\vartheta,n} + Y_{\vartheta,n}$ converge to X in distribution uniformly over Θ.*

Proof. For $\delta > 0$ and any set $B \subseteq R^d$ we define

$$\begin{aligned} B^{\delta} &:= \{y \in \mathbb{R}^d \mid |x-y| < \delta \text{ for some } x \in B\}, \\ B^{-\delta} &:= \{y \in \mathbb{R}^d \,|x \in B \text{ for all } x \text{ with } |x-y| < \delta\} \\ &= \mathbb{R}^d \setminus \{y \in \mathbb{R}^d \mid \text{there is an } x \in \mathbb{R}^d \setminus B \text{ such that } |x-y| < \delta\}. \end{aligned}$$

As B^{δ} and $B^{-\delta}$ are open and closed since they are the union of open balls and the complement of a union of open balls, respectively. Consequently, they are Borel sets. It holds $\bigcap_{\delta>0} B^{\delta} = \overline{B}$ and $\bigcup_{\delta>0} B^{-\delta} = B^{\circ}$. For B with $\mathbb{P}(X \in \partial B) = 0$, we have $\mathbb{P}(X \in B^{\circ}) = \mathbb{P}(X \in \overline{B}) = \mathbb{P}(X \in B)$. Consequently

$$\lim_{\delta\to 0} \mathbb{P}(X \in B^{\delta}) = \lim_{\delta\to 0} \mathbb{P}(X \in B^{-\delta}) = \mathbb{P}(X \in B). \tag{5.2}$$

Let $\eta > 0$ be given. For all $\delta > 0$ it holds

$$\begin{aligned} &\sup_{\vartheta\in\Theta} \mathbb{P}(Z_{\vartheta,n}X_{\vartheta,n} + Y_{\vartheta,n} \in B) \\ &\leqslant \sup_{\vartheta\in\Theta} \mathbb{P}(Z_{\vartheta,n}X_{\vartheta,n} \in B^{\delta}) + \sup_{\vartheta\in\Theta} \mathbb{P}(|Y_{\vartheta,n}| \geqslant \delta), \end{aligned}$$

for large n the second term is smaller than η owing to the assumptions on $Y_{\vartheta,n}$,

$$\leqslant \sup_{\vartheta\in\Theta} \mathbb{P}(Z_{\vartheta,n}X_{\vartheta,n} \in B^{\delta}) + \eta.$$

As a single random vector X is tight, meaning that for each $\eta > 0$ there is M such that $\mathbb{P}(|X| \geqslant M) < \eta$. By taking the set $\{x \in \mathbb{R}^d \mid |x| \geqslant M\}$ in the definition of uniform convergence in total variation, we obtain for n large enough $\sup_{\vartheta\in\Theta} \mathbb{P}(|X_{\vartheta,n}| \geqslant M) < 2\eta$. By considering possibly larger n we can also ensure $\sup_{\vartheta\in\Theta} \mathbb{P}(|Z_{\vartheta,n} - 1| \geqslant \delta/M) \leqslant \eta$. But if $|X_{\vartheta,n}| < M$ and $|Z_{\vartheta,n} - 1| < \delta/M$, then $|X_{\vartheta,n}Z_{\vartheta,n} - X_{\vartheta,n}| < \delta$. For n large enough we have

$$\begin{aligned} \sup_{\vartheta\in\Theta} \mathbb{P}(Z_{\vartheta,n}X_{\vartheta,n} + Y_{\vartheta,n} \in B) &\leqslant \sup_{\vartheta\in\Theta} \mathbb{P}(Z_{\vartheta,n}X_{\vartheta,n} \in B^{\delta}) + \eta \\ &\leqslant \sup_{\vartheta\in\Theta} \mathbb{P}(X_{\vartheta,n} \in B^{2\delta}) + 4\eta \end{aligned}$$

$$\begin{aligned}&\leqslant \mathbb{P}(X \in B^{2\delta}) + 5\eta \\ &\leqslant \mathbb{P}(X \in B) + 6\eta, \end{aligned} \tag{5.3}$$

for δ small enough by (5.2).

Let $\eta > 0$ be given. For n large enough and $\delta > 0$ small enough we obtain similarly

$$\begin{aligned}\inf_{\vartheta\in\Theta} \mathbb{P}(Z_{\vartheta,n}X_{\vartheta,n} + Y_{\vartheta,n} \in B) &\geqslant \inf_{\vartheta\in\Theta} \mathbb{P}(Z_{\vartheta,n}X_{\vartheta,n} \in B^{-\delta}) - \eta \\ &\geqslant \inf_{\vartheta\in\Theta} \mathbb{P}(X_{\vartheta,n} \in B^{-2\delta}) - 4\eta \\ &\geqslant \mathbb{P}(X \in B^{-2\delta}) - 5\eta \\ &\geqslant \mathbb{P}(X \in B) - 6\eta. \end{aligned} \tag{5.4}$$

The statement follows by combining (5.3) and (5.4). □

Lemma 5.4 outlines how to proceed in showing uniform convergence. $X_{\vartheta,n}$ will be the leading term of the linearized stochastic error, $Y_{\vartheta,n}$ will be the sum of the smaller stochastic errors, the remainder term and the approximation error, while $Z_{\vartheta,n}$ will be the quotient of standard deviation and estimated standard deviation. The approximation error is uniformly controlled over $\mathcal{G}_s(R, \sigma_{\max})$ and thus over $\mathcal{H}_s(R, \sigma_{\max})$, too.

The substitution of the standard deviation by its empirical counterpart works as follows. We fix some x and write d instead of $d(x)$ in Theorem 4.1 to unify the notation with Theorem 4.4 and to treat both simultaneously. For δ continuous and positive the standard deviation depends continuously on γ and λ through d. γ and λ are restricted to a compact set. By the uniform convergence and by an upper bound of the standard deviation we obtain $\sup_{\mathcal{T}} \mathbb{P}_{\mathcal{T}}(|\Delta\hat{\rho}| > \kappa) \to 0$ for all $\kappa > 0$ and for $\rho \in \{\gamma, \lambda\}$, where the supremum is over all $\mathcal{T} \in \mathcal{H}_s(R, \sigma_{\max})$ with a fixed volatility σ. Since d is uniformly continuous in γ and λ, we obtain $\sup_{\mathcal{T}} \mathbb{P}_{\mathcal{T}}(|\Delta\hat{d}| > \kappa) \to 0$ for all $\kappa > 0$, which gives the assumption on $d/\hat{d}$ corresponding to $Z_{\vartheta,n}$ in Lemma 5.4 by a lower bound on d.

The following lemma shows that uniform convergence in total variation of the linearized stochastic error follows from uniform convergence in each component of the covariance matrix.

Lemma 5.5. *Let X be a normal random vector with symmetric positive definite covariance matrix $A \in \mathbb{R}^{d\times d}$. Let $X_{\vartheta,n}$, $n \in \mathbb{N}$, $\vartheta \in \Theta$, be normal random vectors with covariance matrices $A_{\vartheta,n} \in \mathbb{R}^{d\times d}$. If $A_{\vartheta,n}$ converge to A in each component uniformly over Θ as $n \to \infty$, then $X_{\vartheta,n}$ converge to X in total variation uniformly over Θ as $n \to \infty$.*

Proof. We have to show that

$$\sup_{\vartheta\in\Theta} \sup_{B} \left| \int_B \frac{\exp(-\langle x, A_{\vartheta,n}^{-1}x\rangle/2)}{\det(\sqrt{2\pi}A_{\vartheta,n})} - \frac{\exp(-\langle x, A^{-1}x\rangle/2)}{\det(\sqrt{2\pi}A)} \mathrm{d}x \right| \tag{5.5}$$

converges to zero as $n \to \infty$. The determinant is a continuous function of the components of a matrix. For all $\delta \in (0, \det(\sqrt{2\pi}A))$ there is $N \in \mathbb{N}$ such that $\sup_{\vartheta\in\Theta} |\det(\sqrt{2\pi}A_{\vartheta,n}) - \det(\sqrt{2\pi}A)| \leqslant \delta$ holds for all $n \geqslant N$ and $A_{\vartheta,n}^{-1}$ is well–defined for all $\vartheta \in \Theta$ if $n \geqslant N$. Expression (5.5) equals

$$\begin{aligned}
&= \sup_{\vartheta\in\Theta} \frac{1}{2} \int_{\mathbb{R}^d} \left| \frac{\exp(-\langle x, A_{\vartheta,n}^{-1}x\rangle/2)}{\det(\sqrt{2\pi}A_{\vartheta,n})} - \frac{\exp(-\langle x, A^{-1}x\rangle/2)}{\det(\sqrt{2\pi}A)} \right| \mathrm{d}x \\
&= \sup_{\vartheta\in\Theta} \frac{1}{2} \int_{\mathbb{R}^d} \frac{\exp(-\langle x, A^{-1}x\rangle/2)}{|\det(\sqrt{2\pi}A)|} \\
&\qquad \left| \frac{\det(\sqrt{2\pi}A)\exp(-\langle x, (A_{\vartheta,n}^{-1} - A^{-1})x\rangle/2)}{\det(\sqrt{2\pi}A_{\vartheta,n})} - 1 \right| \mathrm{d}x. \qquad (5.6)
\end{aligned}$$

$A_{\vartheta,n}$ converges to A in each component uniformly over Θ. Likewise $A_{\vartheta,n}^{-1}$ converges to A^{-1} in each component uniformly over Θ. We now additionally require $\delta \in (0, \lambda_{\min}(A^{-1})/d)$, where $\lambda_{\min}(A^{-1})$ is the smallest eigenvalue of A^{-1}. By going over to a possibly larger N, we may assume that $\sup_{\vartheta\in\Theta} |(A_{\vartheta,n}^{-1} - A^{-1})_{jk}| \leqslant \delta$ for all $j, k = 1, \dots, d$, for all $n \geqslant N$. Then for all $n \geqslant N$

$$\begin{aligned}
|\langle x, (A_{\vartheta,n}^{-1} - A^{-1})x\rangle| &= \left| \sum_{j,k=1}^{d} x_j (A_{\vartheta,n}^{-1} - A^{-1})_{jk} x_k \right| \\
&\leqslant \delta \left(\sum_{j=1}^{d} |x_j| \right)^2 \leqslant \delta d \|x\|_2^2,
\end{aligned}$$

where in the last step the Cauchy–Schwarz inequality is used. We see that (5.6) can be bounded by

$$\begin{aligned}
&\frac{1}{2} \int_{\mathbb{R}^d} \frac{\exp(-\langle x, A^{-1}x\rangle/2)}{|\det(\sqrt{2\pi}A)|} \\
&\left(\frac{|\det(\sqrt{2\pi}A)| \exp(d\delta\|x\|_2^2/2)}{|\det(\sqrt{2\pi}A)| - \delta} - \frac{|\det(\sqrt{2\pi}A)| \exp(-d\delta\|x\|_2^2/2)}{|\det(\sqrt{2\pi}A)| + \delta} \right) \mathrm{d}x. \qquad (5.7)
\end{aligned}$$

The integrand converges pointwise to zero for $\delta \to 0$. For fixed δ the function is at the same time a dominating function, which is integrable since $A^{-1} - \delta d I_d$ is positive definite by the choice $\delta \in (0, \lambda_{\min}(A^{-1})/d)$. By the dominated convergence theorem, (5.7) converges to zero and likewise (5.5). □

We will show uniform convergence in each component of the covariance matrix. By Lemma 5.5 this leads to uniform convergence in total variation of the linearized stochastic errors provided that the covariance matrix of the limit is nondegenerate. To this end, we assume in the case $\sigma = 0$ that δ is positive and that $\int_0^1 u^4 w_\rho^1(u)^2 \mathrm{d}u > 0$ for the weight functions w_ρ^1, $\rho \in \{\sigma, \gamma, \lambda, 0, \mu\}$, involved. Joint distributions may further only involve more than one of the estimators $\hat{\sigma}^2$, $\hat{\lambda}$ or $\hat{\mu}(0)$, if the covariance matrix is positive definite. In the case $\sigma > 0$, we assume that $w_\sigma^1(1), w_\gamma^1(1), w_\lambda^1(1), w_\mu^1(1) \neq 0$ and $\|\delta^2\|_{L^2(\mathbb{R})} > 0$. By the uniform version of Lemma 4.10 the remainder term converges uniformly to zero. By Lemma 5.4 each of the rescaled stochastic error terms

$$\begin{aligned}
&\frac{1}{\epsilon_n \exp(T\sigma^2 U^2/2) w_\sigma^1(1)} 2 \int_0^1 \mathrm{Re}(\Delta\psi_n(Uu)) w_\sigma^1(u) \mathrm{d}u, \\
&\frac{1}{\epsilon_n \exp(T\sigma^2 U^2/2) w_\gamma^1(1)} 2 \int_0^1 \mathrm{Im}(\Delta\psi_n(Uu)) w_\gamma^1(u) \mathrm{d}u, \\
&\frac{1}{\epsilon_n \exp(T\sigma^2 U^2/2) w_\lambda^1(1)} 2 \int_0^1 \mathrm{Re}(\Delta\psi_n(Uu)) w_\lambda^1(u) \mathrm{d}u, \\
&\frac{2\pi}{\epsilon_n \exp(T\sigma^2 U^2/2) w_\mu^1(1)} \mathcal{F}^{-1}\left[\Delta\psi_n(Uu) w_\mu^1(u)\right](Ux)
\end{aligned}$$

converges uniformly in distribution to

$$\frac{\sqrt{2}\|\delta^2\|_{L^2(\mathbb{R})}}{\exp(T(\sigma^2/2 + \gamma - \lambda)) T^2 \sigma^2} W,$$

where W is a standard normal random variable. By the uniform versions of Lemma 4.8 and Lemma 4.10 we also obtain that for the second and third of the above stochastic error terms holds joint uniform convergence in distribution to independent normal variables.

5.2 Uniformity in the case $\sigma = 0$

The convergence in distribution of the linearized stochastic errors is shown in Lemma 4.5 by the convergence of the components in the covariance matrix. We restrict ourselves to the case $x_1 = \cdots = x_n$ and show uniform convergence in each component of the covariance matrix. This implies uniform convergence in total variation by Lemma 5.5. We assume that δ is continuous at all $x \in [x_1 - TR, x_1 + TR]$. We note that f_U, g_U and θ_j depend on the characteristic triplet $\mathcal{T}$. The uniform convergence of the rescaled covariances (4.11) will be shown by the following easy lemma.

Lemma 5.6. *Let $f_{\vartheta,n}, f_\vartheta, g_{\vartheta,n}, g_\vartheta \in L^1([0,1],\mathbb{C})$, $n \in \mathbb{N}$, $\vartheta \in \Theta$, and $M > 0$ be such that $\|f_{\vartheta,n}\|_\infty, \|g_\vartheta\|_\infty \leqslant M$ for all $n \in \mathbb{N}$, for all $\vartheta \in \Theta$. Let $\sup_{\vartheta\in\Theta} \|f_{\vartheta,n} - f_\vartheta\|_{L^1([0,1],\mathbb{C})} \to 0$ and $\sup_{\vartheta\in\Theta} \|g_{\vartheta,n} - g_\vartheta\|_{L^1([0,1],\mathbb{C})} \to 0$ as $n \to \infty$. Then*

$$\sup_{\vartheta\in\Theta} \int_0^1 |f_{\vartheta,n}(x) g_{\vartheta,n}(x) - f_\vartheta(x) g_\vartheta(x)| \mathrm{d}x \to 0 \qquad \text{as } n \to \infty.$$

Proof. For all $\vartheta \in \Theta$ it holds

$$\begin{aligned}
|f_{\vartheta,n} g_{\vartheta,n} - f_\vartheta g_\vartheta| &\leqslant |f_{\vartheta,n} g_{\vartheta,n} - f_{\vartheta,n} g_\vartheta| + |f_{\vartheta,n} g_\vartheta - f_\vartheta g_\vartheta| \\
&\leqslant M|g_{\vartheta,n} - g_\vartheta| + M|f_{\vartheta,n} - f_\vartheta|.
\end{aligned}$$

By the assumptions $\sup_{\vartheta\in\Theta} \|f_{\vartheta,n} - f_\vartheta\|_{L^1([0,1],\mathbb{C})} \to 0$ as $n \to \infty$ and $\sup_{\vartheta\in\Theta} \|g_{\vartheta,n} - g_\vartheta\|_{L^1([0,1],\mathbb{C})} \to 0$ as $n \to \infty$ and the claimed statement follows. $\square$

Let us verify the assumptions of Lemma 5.6 for the first factor in (4.11). f_U and $u^2 w_j(u)$ will correspond to $f_{\vartheta,n}$ and f_ϑ, respectively. It holds

$$\begin{aligned}
&|f_U(u) + u^2 w_j(u)| \\
&= |u^2 w_j(u)(1 - \exp(-T\mathcal{F}\mu(Uu))) + iu \exp(-T\mathcal{F}\mu(Uu))/U| \\
&\leqslant u^2 |w_j(u)|(\exp(TR(1 \wedge h(Uu))) - 1) + iu \exp(TR)/U.
\end{aligned} \tag{5.8}$$

This bound does not depend on $\mathcal{T}$ and converges everywhere to zero. Further the bound $f_U(u), u^2 w_j(u) \leqslant \sqrt{2}\|w_j\|_\infty \exp(TR)$ does not depend on U nor on $\mathcal{T}$. By the dominated convergence theorem

$$\sup_{\mathcal{T}} \|f_U(u) + u^2 w_j(u)\|_{L^1([0,1],\mathbb{C})} \to 0 \tag{5.9}$$

as $U \to \infty$ and the conditions on the first factor in Lemma 5.6 are satisfied.

To show the assumptions of Lemma 5.6 on the second factor, which is the complex conjugate of

$$\begin{aligned}
&(g_U(v) * \mathcal{F}^{-1}(\delta(y/U + \theta_j)^2)(v))(u) \\
&= \int_{-\infty}^{\infty} g_U(u - v)\mathcal{F}^{-1}(\delta(y + \theta_j)^2)(Uv) U \mathrm{d}v,
\end{aligned}$$

we apply the following lemma. It is a uniform version of the basic theorem on approximate identities, compare Theorem 1.2.21 in Grafakos (2004).

Lemma 5.7. *Let $f, f_{\vartheta,n} \in L^1(\mathbb{R},\mathbb{C})$, $n \in \mathbb{N}$, $\vartheta \in \Theta$. Let $\sup_{\vartheta\in\Theta} \|f_{\vartheta,n} - f\|_{L^1(\mathbb{R},\mathbb{C})} \to 0$ for $n \to \infty$. Let $\delta_{\vartheta,n} \in L^1(\mathbb{R},\mathbb{C})$ fulfill the following properties:*

(i) There exists $c > 0$ such that $\|\delta_{\vartheta,n}\|_{L^1(\mathbb{R},\mathbb{C})} \leqslant c$ for all $n \in \mathbb{N}$, for all $\vartheta \in \Theta$.

(ii) $\int_{-\infty}^{\infty} \delta_{\vartheta,n}(y)\mathrm{d}y = c_\vartheta$ for all $n \in \mathbb{N}$, for all $\vartheta \in \Theta$.

(iii) For any neighborhood V of zero we have $\sup_{\vartheta\in\Theta} \int_{V^c} |\delta_{\vartheta,n}(y)|\mathrm{d}y \to 0$ as $n \to \infty$.

Then $\sup_{\vartheta\in\Theta} \|\delta_{\vartheta,n} * f_{\vartheta,n} - c_\vartheta f\|_{L^1(\mathbb{R},\mathbb{C})} \to 0$ as $n \to \infty$.

Proof. By the triangle inequality

$$\begin{aligned}
&\sup_{\vartheta\in\Theta} \|\delta_{\vartheta,n} * f_{\vartheta,n} - c_\vartheta f\|_{L^1(\mathbb{R},\mathbb{C})} \\
&\leqslant \sup_{\vartheta\in\Theta} \|\delta_{\vartheta,n} * (f_{\vartheta,n} - f)\|_{L^1(\mathbb{R},\mathbb{C})} + \sup_{\vartheta\in\Theta} \|\delta_{\vartheta,n} * f - c_\vartheta f\|_{L^1(\mathbb{R},\mathbb{C})}.
\end{aligned} \tag{5.10}$$

The first term in the triangle inequality (5.10) can be bounded by

$$\begin{aligned}\sup_{\vartheta\in\Theta}\|\delta_{\vartheta,n}*(f_{\vartheta,n}-f)\|_{L^1(\mathbb{R},\mathbb{C})} &\leqslant \sup_{\vartheta\in\Theta}\|\delta_{\vartheta,n}\|_{L^1(\mathbb{R},\mathbb{C})}\,\|f_{\vartheta,n}-f\|_{L^1(\mathbb{R},\mathbb{C})}\\ &\leqslant c\,\sup_{\vartheta\in\Theta}\|f_{\vartheta,n}-f\|_{L^1(\mathbb{R},\mathbb{C})},\end{aligned}$$

which converges to zero by assumption. To bound the second term in the triangle inequality (5.10) proceed as in the proof of Theorem 1.2.19 in (Grafakos, 2004, p. 26). Without loss of generality we may assume that $\|f\|_{L^1(\mathbb{R},\mathbb{C})} > 0$. Continuous functions with compact support are dense in $L^1(\mathbb{R},\mathbb{C})$. Since continuous functions g with compact support are bounded we obtain by the dominated convergence theorem

$$\int_{-\infty}^{\infty}|g(x-y)-g(x)|\mathrm{d}x\to 0$$

as $y\to 0$. We approximate $f\in L^1(\mathbb{R},\mathbb{C})$ by a continuous function with compact support and see that for all $\delta>0$ there is some neighborhood V of zero such that

$$\int_{-\infty}^{\infty}|f(x-y)-f(x)|\mathrm{d}x<\frac{\delta}{2c}\qquad\text{for all } y\in V. \tag{5.11}$$

It further holds

$$\begin{aligned}&(\delta_{\vartheta,n}*f)(x)-c_\vartheta f(x)\\ &=(\delta_{\vartheta,n}*f)(x)-f(x)\int_{-\infty}^{\infty}\delta_{\vartheta,n}(y)\mathrm{d}y\\ &=\int_{-\infty}^{\infty}(f(x-y)-f(x))\delta_{\vartheta,n}(y)\mathrm{d}y\\ &=\int_V(f(x-y)-f(x))\delta_{\vartheta,n}(y)\mathrm{d}y+\int_{V^c}(f(x-y)-f(x))\delta_{\vartheta,n}(y)\mathrm{d}y.\end{aligned}$$

We take L^1–norms with respect to x and obtain for the first term by (5.11)

$$\begin{aligned}&\sup_{\vartheta\in\Theta}\left\|\int_V(f(x-y)-f(x))\delta_{\vartheta,n}(y)\mathrm{d}y\right\|_{L^1(\mathbb{R},\mathbb{C})}\\ &\leqslant\sup_{\vartheta\in\Theta}\int_V\|f(x-y)-f(x)\|_{L^1(\mathbb{R},\mathbb{C})}|\delta_{\vartheta,n}(y)|\mathrm{d}y\\ &\leqslant\sup_{\vartheta\in\Theta}\int_V\frac{\delta}{2c}|\delta_{\vartheta,n}(y)|\mathrm{d}y<\frac{\delta}{2}\end{aligned}\tag{5.12}$$

and for the second term

$$\sup_{\vartheta\in\Theta}\left\|\int_{V^c}(f(x-y)-f(x))\delta_{\vartheta,n}(y)\mathrm{d}y\right\|_{L^1(\mathbb{R},\mathbb{C})}$$

$$\leqslant \sup_{\vartheta\in\Theta} \int_{V^c} 2\|f\|_{L^1(\mathbb{R},\mathbb{C})} |\delta_{\vartheta,n}(y)| \mathrm{d}y < \frac{\delta}{2}, \tag{5.13}$$

where n is taken large enough such that

$$\sup_{\vartheta\in\Theta} \int_{V^c} |\delta_{\vartheta,n}(y)| \mathrm{d}y < \frac{\delta}{4\|f\|_{L^1(\mathbb{R},\mathbb{C})}}.$$

The lemma is a consequence of (5.12) and (5.13). □

Let us first verify the assumptions (i), (ii) and (iii) for $\mathcal{F}^{-1}(\delta(y+\theta_j)^2(Uv)U$, which will correspond to $\delta_{\vartheta,n}$. We have

$$\mathcal{F}^{-1}(\delta(y+\theta)^2)(v) = e^{i\theta v}\mathcal{F}^{-1}(\delta(y)^2)(v). \tag{5.14}$$

By the assumptions of Lemma 4.5 it holds $\mathcal{F}\delta^2 \in L^1(\mathbb{R})$. The equality

$$|\mathcal{F}^{-1}(\delta(y+\theta_j)^2)(Uv)U| = |\mathcal{F}^{-1}(\delta(y)^2)(Uv)U|$$

shows that the absolute value does not depend on $\mathcal{T}$ and that conditions (i) and (iii) are satisfied. Condition (ii) is satisfied by (4.12). We have

$$\sup_{\mathcal{T}} \|g_U(u) + u^2 w_k(u)\|_{L^1([0,1],\mathbb{C})} \to 0 \tag{5.15}$$

as in the corresponding equation (5.9) for f_U. By extending g_U and w_k by zero outside $[0,1]$ this holds in $L^1(\mathbb{R},\mathbb{C})$, too. We apply Lemma 5.7 to $g_U(v)$ and $v^2 w_k(v)$, which correspond to $f_{\vartheta,n}$ and f in this lemma. $v^2 w_k(v)$ is bounded and since condition (i) is satisfied, we see that $\delta(\theta_j)^2 v^2 w_k(v)$ is uniformly bounded over all characteristic triplets. By Lemma 5.6 the convergence of the covariances (4.13) holds uniformly. The convergence in the analogous equation without conjugation (4.14) holds uniformly, too.

5.3 Uniformity in the case $\sigma > 0$

Let us first fix $\sigma > 0$ and prove uniform convergence for all characteristic triplets with this fixed value of σ. To this end, we show uniform convergence in Lemma 4.6. In order to control the error when going over to smaller domain of integration in (4.21) the term $\overline{\mathcal{F}^{-1}(\delta(y+\theta)^2)(U(u-v))} f_U(u)\overline{g_U(v)}/(uv)$ needs to be bounded uniformly. The inverse Fourier transform is bounded by the L^1–norm of δ^2. The functions f_U and g_U are uniformly bounded since $\|\mathcal{F}\mu\|_\infty \leqslant R$. The crucial step is the limit (4.22), where the refined dirac sequence argument is applied. As stated in (5.14), a translation before the Fourier transform is equal to a multiplication by a complex unit after the Fourier transform. Since $|\theta| \leqslant T(\sigma^2_{\max}+R)$, the complex unit tends uniformly to one. w_U and $\widetilde{w}_U$ do not depend on the characteristic triplet. Since there is h with $|\mathcal{F}\mu(u)| \leqslant R\,h(u)$ and $h(u) \to 0$ as $|u| \to \infty$, the factor $(-u+i/U)/\exp(T\mathcal{F}\mu(Uu))$ converges to $-u$ uniformly over $\mathcal{H}_s(R,\sigma_{\max})$. This leads to uniform convergence in (4.22) and thus in (4.23). To see that the covariance without conjugation converges in (4.24) uniformly to zero, we observe that $\mathcal{F}^{-1}(\delta(y+\theta)^2)(u) \to 0$ for $|u| \to \infty$ uniformly since the translation by θ corresponds to a multiplication by a complex unit by equation (5.14). We immediately obtain the uniform convergence in each component of the covariance matrix in Lemma 4.8.

5.4 Uniformity for the remainder term

This section treats uniform versions of Proposition 3.3, Lemma 4.10 and Lemma 4.11. We begin with the uniform version of Proposition 3.3. More precisely, for $q \geqslant 1$, $R \geqslant 0$ and $\sigma_0 > 0$ there exist $M > 0$ and $x_0 \in \mathbb{R}$ such that for all characteristic triplets in $\mathcal{H}_s(R, \sigma_0)$ and for all $U \geqslant x_0$

$$\mathbb{E}\left[\sup_{u\in[-1,1]} |\mathcal{L}_{n,U}(u)|^q\right]^{1/q} \leqslant M\epsilon_n U^2 \exp(T\sigma_0^2 U^2/2).$$

In order to see this, we observe that $X(u) = \int_{-\infty}^{\infty} e^{iux}\delta(x)\mathrm{d}W(x)$ does not depend on the characteristic triplet. Let us first consider the characteristic triplets in $\mathcal{H}_s(R, \sigma_0)$ with $\sigma^2 \in [0, \sigma_0^2/2]$. The uniformity in the characteristic triplets is a consequence of (3.6) and $\sqrt{\log(U)}\exp(T\sigma^2U^2/2) = O(\exp(T\sigma_0^2U^2/2))$ both holding uniformly for $\sigma^2 \in [0, \sigma_0^2/2]$. For $\sigma^2 \in (\sigma_0^2/2, \sigma_0^2]$ we proceed as for $\sigma > 0$ in the proof of Proposition 3.3. Here the key observation is that (3.7) holds uniformly for $\sigma^2 \in (\sigma_0^2/2, \sigma_0^2]$. The similar statement that for $q \geqslant 1$ and $R \geqslant 0$ there exist $M > 0$ and $x_0 \in \mathbb{R}$ such that for all characteristic triplets in $\mathcal{H}_s(R, 0)$ and for all $U \geqslant x_0$

$$\mathbb{E}\left[\sup_{u\in[-1,1]} |\mathcal{L}_{n,U}(u)|^q\right]^{1/q} \leqslant M\epsilon_n U^2\sqrt{\log(U)}$$

follows directly from the uniformity of (3.6) in the characteristic triplets.

Lemma 4.10 and Lemma 4.11 hold uniformly, meaning that for all $\eta > 0$

$$\sup_{\mathcal{T}\in\mathcal{H}_s(R,\sigma_0)} \mathbb{P}_{\mathcal{T}}\left(\left|\frac{1}{\epsilon_n\exp(T\sigma_0^2U_n^2/2)}\int_0^1 w_{U_n}(u)\mathcal{R}_{n,U_n}(u)\mathrm{d}u\right| > \eta\right) \to 0,$$
$$\sup_{\mathcal{T}\in\mathcal{H}_s(R,0)} \mathbb{P}_{\mathcal{T}}\left(\left|\frac{1}{\epsilon_n U_n^{3/2}}\int_0^1 w_{U_n}(u)\mathcal{R}_{n,U_n}(u)\mathrm{d}u\right| > \eta\right) \to 0,$$

as $n \to \infty$. For Lemma 4.11 this follows from the uniform convergence of the bound in the proof. For Lemma 4.10 this can be seen by the corresponding uniform statements along the lines of the proof up to the bound (4.28). Then we bound (4.28) in two different ways depending on whether $\sigma^2 \in [0, \sigma_0^2/2]$ or $\sigma^2 \in (\sigma_0^2/2, \sigma_0^2]$. In the latter case we can proceed as in the proof for $\sigma > 0$. Since σ^2 is bounded from below by $\sigma_0^2/2 > 0$ the convergence is uniform. For $\sigma^2 \in [0, \sigma_0^2/2]$ we estimate

$$\begin{aligned}
&\frac{C}{\epsilon_n \exp(T\sigma_0^2 U^2/2)} \int_0^1 \frac{\epsilon_n^2 (U^2+U^4) u \exp(T\sigma^2 U^2 u^2) \|\delta\|_{L^2(\mathbb{R})}^2}{T^2 \exp(2T(\sigma^2/2+\gamma-\lambda) - 2T\|\mathcal{F}\mu\|_\infty)} \mathrm{d}u \\
&\leqslant \frac{C\epsilon_n (U^2+U^4) \|\delta\|_{L^2(\mathbb{R})}^2}{T^2 \exp(2T(\sigma^2/2+\gamma-\lambda) - 2T\|\mathcal{F}\mu\|_\infty)},
\end{aligned}$$

which converges uniformly to zero by $\sigma_0 > 0$ and by the assumption

$$\epsilon_n U^2 \exp(T\sigma_0^2 U^2/2) \to 0$$

of Lemma 4.10. The maximum of the bounds is a bound that holds for all characteristic triplets in the class. This shows the uniform version of Lemma 4.10.

6 Applications

In this chapter, we construct confidence intervals and joint confidence sets for the spectral calibration method. The construction is based on the asymptotic normality results in Chapter 4 and on the uniform versions in Chapter 5. Thanks to the uniformity with respect to the underlying probability measure the confidence intervals are honest. Special attention is paid to joint confidence sets if the covariance of two estimation errors does not converge. Inference for the volatility is treated in a separate section. Since for the choice of the cut–off value the volatility itself is used, it is not clear how to construct confidence intervals for the volatility directly. We circumvent this problem by a testing approach and by choosing the cut–off value according to the value of the volatility given by the null hypothesis. Varying the null hypothesis in an interval yields a family of tests, from which we can construct a confidence set for the volatility. So the confidence set for the volatility is obtained only indirectly with the intermediate step of constructing a family of tests on the value of the volatility. In Section 6.1, we consider confidence intervals and sets for the drift, the intensity and pointwise for the jump density. The inference on the volatility is treated in Section 6.2.

6.1 Construction of confidence intervals and confidence sets

For $\sigma = 0$ we define confidence intervals

$$\begin{aligned} I_{\gamma,n} &:= \ [\hat{\gamma} - \hat{s}_\gamma \epsilon_n U_n^{1/2} q_{\eta/2}, & \hat{\gamma} + \hat{s}_\gamma \epsilon_n U_n^{1/2} q_{\eta/2}], \\ I_{\lambda,n} &:= \ [\hat{\lambda} - \hat{s}_\lambda \epsilon_n U_n^{3/2} q_{\eta/2}, & \hat{\lambda} + \hat{s}_\lambda \epsilon_n U_n^{3/2} q_{\eta/2}], \\ I_{\mu(0),n} &:= \ [\hat{\mu}(0) - \hat{s}_{\mu(0)} \epsilon_n U_n^{5/2} q_{\eta/2}, & \hat{\mu}(0) + \hat{s}_{\mu(0)} \epsilon_n U_n^{5/2} q_{\eta/2}], \\ I_{\mu(x),n} &:= \ [\hat{\mu}(x) - \hat{s}_{\mu(x)} \epsilon_n U_n^{5/2} q_{\eta/2}, & \hat{\mu}(x) + \hat{s}_{\mu(x)} \epsilon_n U_n^{5/2} q_{\eta/2}], \end{aligned} \tag{6.1}$$

where $x \in \mathbb{R}\backslash\{0\}$, $\eta \in (0,1)$, q_η denotes the $(1-\eta)$–quantile of the standard normal distribution and

$$\begin{pmatrix} \hat{s}_\gamma \\ \hat{s}_\lambda \\ \hat{s}_{\mu(0)} \\ \hat{s}_{\mu(x)} \end{pmatrix} := \begin{pmatrix} \hat{s}(0) \left(\int_0^1 u^4 w_\gamma^1(u)^2 \mathrm{d}u \right)^{1/2} \\ \hat{s}(0) \left(\int_0^1 u^4 w_\lambda^1(u)^2 \mathrm{d}u \right)^{1/2} \\ \hat{s}(0) \left(\int_0^1 u^4 w_0(u)^2 \mathrm{d}u \right)^{1/2} /(2\pi) \\ \hat{s}(x) \left(\int_0^1 u^4 w_\mu^1(u)^2 \mathrm{d}u \right)^{1/2} /(2\pi) \end{pmatrix}$$

with $\hat{s}(x) := 2\sqrt{\pi}\delta(x + T\hat{\gamma})\exp(T(\hat{\lambda} - \hat{\gamma}))/T$. We fix some arbitrarily slowly decreasing function h with $h(u) \to 0$ as $|u| \to \infty$. We denote by $\mathcal{H}_s(R, \sigma_{\max})$ the subset of characteristic triplets in $\mathcal{G}_s(R, \sigma_{\max})$ that satisfy the additional conditions (5.1), which are

$$\|\mathcal{F}\mu\|_\infty \leqslant R \quad \text{and} \quad |\mathcal{F}\mu(u)| \leqslant R\,h(u), \quad \forall u \in \mathbb{R}\,.$$

The additional conditions are used to extend the convergence in the theorems to be uniform over all characteristic triplets in $\mathcal{H}_s(R, \sigma_{\max})$, see Theorem 5.1, and to obtain *honest* confidence sets meaning that the level is achieved uniformly over a class of characteristic triplets.

Corollary 6.1. *Let $\sigma = 0$. On the assumptions of Theorem 4.1 and on the assumption that δ is positive and continuous*

$$\lim_{n\to\infty} \inf_{\mathcal{T}\in\mathcal{H}_s(R,0)} \mathbb{P}_\mathcal{T}(\rho \in I_{\rho,n}) = 1 - \eta$$

holds for the intervals (6.1) *and for all $\rho \in \{\gamma, \lambda, \mu(x) | x \in \mathbb{R}\}$.*

If the infimum in the corollary is omitted, then the statement holds for all characteristic triplets $\mathcal{T}$ in $\mathcal{G}_s(R, 0)$ and is then a direct consequence of Theorem 4.1. The same holds for the other confidence intervals and sets, where in the case of positive volatility the statements hold for the corresponding characteristic triplets $\mathcal{T}$ in $\mathcal{G}_s(R, \sigma_{\max})$ and follow from Theorem 4.4.

Remark 6.2. For the two parameters γ and λ we define the confidence set $A_n := \{(\hat{\gamma} + \epsilon_n U^{1/2}\hat{s}_\gamma x, \hat{\lambda} + \epsilon_n U^{3/2}\hat{s}_\lambda y)^\top |\, x^2 + y^2 \leqslant k_\eta\}$, where k_η denotes the $(1-\eta)$–quantile of the chi–squared distribution χ_2^2 with two degrees of freedom. Then it holds

$$\lim_{n\to\infty} \inf_{\mathcal{T}\in\mathcal{H}_s(R,0)} \mathbb{P}_\mathcal{T}((\gamma, \lambda)^\top \in A_n) = 1 - \eta.$$

For $\sigma > 0$ we define confidence intervals

$$\begin{aligned}
I_{\gamma,n} &:= [\hat{\gamma} - \hat{s}_\gamma \epsilon_n U_n^{-1} e^{T\sigma^2 U_n^2/2} q_{\eta/2}, && \hat{\gamma} + \hat{s}_\gamma \epsilon_n U_n^{-1} e^{T\sigma^2 U_n^2/2} q_{\eta/2}],\\
I_{\lambda,n} &:= [\hat{\lambda} - \hat{s}_\lambda \epsilon_n e^{T\sigma^2 U_n^2/2} q_{\eta/2}, && \hat{\lambda} + \hat{s}_\lambda \epsilon_n e^{T\sigma^2 U_n^2/2} q_{\eta/2}],\\
I_{\mu(0),n} &:= [\hat{\mu}(0) - \hat{s}_{\mu(0)} \epsilon_n U_n e^{T\sigma^2 U_n^2/2} q_{\eta/2}, && \hat{\mu}(0) + \hat{s}_{\mu(0)} \epsilon_n U_n e^{T\sigma^2 U_n^2/2} q_{\eta/2}],\\
I_{\mu(x),n} &:= [\hat{\mu}(x) - \hat{s}_\mu \epsilon_n U_n e^{T\sigma^2 U_n^2/2} q_{\eta/2}, && \hat{\mu}(x) + \hat{s}_\mu \epsilon_n U_n e^{T\sigma^2 U_n^2/2} q_{\eta/2}],
\end{aligned} \tag{6.2}$$

where $x \in \mathbb{R}\setminus\{0\}$,

$$\begin{pmatrix} \hat{s}_\gamma \\ \hat{s}_\lambda \\ \hat{s}_{\mu(0)} \\ \hat{s}_\mu \end{pmatrix} := \frac{\sqrt{2}\|\delta\|_{L^2(\mathbb{R})}}{\exp(T(\sigma^2/2 + \hat{\gamma} - \hat{\lambda}))T^2\sigma^2} \begin{pmatrix} |w_\gamma^1(1)| \\ |w_\lambda^1(1)| \\ |w_0(1)|/(2\pi) \\ |w_\mu^1(1)|/(2\pi) \end{pmatrix}$$

and q_η denotes the $(1-\eta)$–quantile of the standard normal distribution. We assume $\|\delta\|_{L^2(\mathbb{R})} > 0$ and that the weight functions are chosen such that $w_\gamma^1(1)$, $w_\lambda^1(1)$, $w_0(1)$, $w_\mu^1(1) \in \mathbb{R}\setminus\{0\}$. We note that instead of estimating δ nonparametrically it suffices for positive volatility to estimate the L^2–norm of δ. For example, in the case of equidistant design we can first estimate $\mathcal{O}$ with the standard Nadaraya–Watson estimator for regression and then estimate $\|\delta\|_{L^2(\mathbb{R})}$ from the sum of the squared residuals, which leads to a consistent estimator as shown by Dette and Neumeyer (2001).

Corollary 6.3. *Let $\sigma > 0$. On the assumptions of Theorem 4.4*

$$\lim_{n\to\infty} \inf_{\mathcal{T}} \mathbb{P}_{\mathcal{T}}(\rho \in I_{\rho,n}) = 1-\eta$$

holds for the intervals (6.2) *and for all $\rho \in \{\gamma, \lambda, \mu(x) | x \in \mathbb{R}\}$, where the infimum is over all $\mathcal{T} \in \mathcal{H}_s(R, \sigma_{\max})$ with volatility σ.*

For $(\gamma,\lambda)^\top$ a uniform confidence set may be obtained similarly as in the case $\sigma = 0$. Since for $x \in \mathbb{R}\setminus\{0\}$ the covariance of $Z_{n,U_n}(x)$ and V_{n,U_n} and the covariance of $Z_{n,U_n}(x)$ and W_{n,U_n} do not converge, confidence sets for $(\gamma, \mu(x))^\top$ and $(\lambda, \mu(x))^\top$ have to be constructed differently. Let us illustrate how to proceed in this case by constructing a confidence set for $(\mu(x_1), \mu(x_2))^\top$, $x_1, x_2 \in \mathbb{R}\setminus\{0\}$. By Theorem 4.4 the convergence

$$\frac{1}{\epsilon_n \exp(T\sigma^2 U_n^2/2)} \left(\frac{1}{U_n} \begin{pmatrix} \Delta\hat{\mu}(x_1) \\ \Delta\hat{\mu}(x_2) \end{pmatrix} - \frac{d\, w_\mu^1(1)}{2\pi} \begin{pmatrix} Z_{n,U_n}(x_1) \\ Z_{n,U_n}(x_2) \end{pmatrix} \right) \xrightarrow{\mathbb{P}} 0$$

holds for $n \to \infty$. We define

$$M_n := \begin{pmatrix} \cos(U_n x_1) & \sin(U_n x_1) \\ \cos(U_n x_2) & \sin(U_n x_2) \end{pmatrix},$$

and observe that the components of M_n^{-1} are bounded for n for which the absolute value of the determinant is bounded from below by some $c > 0$, i.e., $|\sin(U_n(x_2 - x_1))| \geqslant c$. For such n

$$\frac{1}{\epsilon_n \exp(T\sigma^2 U_n^2/2)} \left(\frac{M_n^{-1}}{U_n} \begin{pmatrix} \Delta\hat{\mu}(x_1) \\ \Delta\hat{\mu}(x_2) \end{pmatrix} - \frac{d\, w_\mu^1(1)}{2\pi} \begin{pmatrix} W_{n,U_n} \\ V_{n,U_n} \end{pmatrix} \right) \xrightarrow{\mathbb{P}} 0$$

holds for $n \to \infty$. We apply the additive version of Slutsky's lemma together with the convergence (4.5) of the appropriately scaled random variables W_{n,U_n} and V_{n,U_n}. In view of the definition of d in (4.4) we observe that $\hat{s}_\mu$ is a consistent estimator of $d|w_\mu^1(1)|/(2\pi)$ and we apply the multiplicative version of Slutsky's lemma, which then leads to

$$\frac{1}{\hat{s}_\mu \epsilon_n U_n \exp(T\sigma^2 U_n^2/2)} M_n^{-1} \begin{pmatrix} \Delta\hat{\mu}(x_1) \\ \Delta\hat{\mu}(x_2) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} W \\ V \end{pmatrix}$$

for $n \to \infty$ such that $|\sin(U_n(x_2 - x_1))| \geqslant c$. We define

$$B_n := (\hat{\mu}(x_1), \hat{\mu}(x_2))^\top + M_n\{\hat{s}_\mu \epsilon_n U_n \exp(T\sigma^2 U_n^2/2)(x,y)^\top \,|\, x^2 + y^2 \leqslant k_\eta\},$$

where k_η denotes the $(1-\eta)$–quantile of the chi–squared distribution χ_2^2 with two degrees of freedom. Then

$$\lim_{\substack{|\sin(U_n(x_2-x_1))|\geqslant c\\ n\to\infty}} \mathbb{P}_\mathcal{T}((\mu(x_1), \mu(x_2))^\top \in B_n) = 1 - \eta$$

holds for all $\mathcal{T} \in \mathcal{G}_s(R, \sigma_{\max}) \cap \{\sigma > 0\}$.

6.2 Inference on the volatility

In this section, we test the hypotheses $H_0 : \mathcal{T} \in \{\sigma = \sigma_0\} \cap \mathcal{H}_s(R, \sigma_{\max})$ against the alternative $H_1 : \mathcal{T} \in \{|\sigma - \sigma_0| \geqslant \tau\} \cap \mathcal{H}_s(R, \sigma_{\max})$ for $\sigma_0 \in [0, \sigma_{\max}]$ and $\tau > 0$. For $\eta \in (0, 1/2]$, the test reaches asymptotically the level η, i.e., $\lim_{n\to\infty} \sup_{\mathcal{T}\in H_0} \mathbb{E}_\mathcal{T}[\varphi_{\sigma_0}] = \eta$. Moreover, the error of the second kind vanishes asymptotically, i.e., $\lim_{n\to\infty} \sup_{\mathcal{T}\in H_1} \mathbb{E}_\mathcal{T}[1 - \varphi_{\sigma_0}] = 0$. This family of tests can be used to construct a confidence set for σ.

For $\sigma_0 > 0$ the most natural test statistic is the following. In order to apply the uniform version of Theorem 4.4 under H_0 we choose a sequence of cut–off values U_n such that $\epsilon_n U_n^2 \exp(T\sigma_0^2 U_n^2/2) \to 0$ and $\epsilon_n U_n^{s+1} \exp(T\sigma_0^2 U_n^2/2) \to \infty$ as $n \to \infty$. Let $\hat{\sigma}^2$ be the estimator corresponding to this cut–off value U_n. Since we assume that $\mathcal{T} \in \mathcal{H}_s(R, \sigma_{\max})$, we have $\sigma \in [0, \sigma_{\max}]$. We ensure $\hat{\sigma}^2 \in [0, \sigma_{\max}^2]$ by taking the maximum with zero and the minimum with $\sigma_{\max}^2$. Likewise we ensure $\hat{\gamma} \in [-R, R]$ and $\hat{\lambda} \in [0, R]$. We assume $\|\delta\|_{L^2(\mathbb{R})} > 0$, choose a weight function with $w_\sigma^1(1) \neq 0$ and define

$$S_{\sigma_0} := \frac{U_n^2(\hat{\sigma}^2 - \sigma_0^2)}{\hat{d}_{\sigma_0}\, \epsilon_n \exp(T\sigma_0^2 U_n^2/2)}, \qquad \hat{d}_{\sigma_0} := \frac{\sqrt{2}\,|w_\sigma^1(1)|\, \|\delta\|_{L^2(\mathbb{R})}}{\exp(T(\sigma_0^2/2 + \hat{\gamma} - \hat{\lambda}))T^2\sigma_0^2}.$$

Under H_0 the test statistic S_{σ_0} converges uniformly in distribution to a standard normal random variable by the uniform version of Theorem 4.4. We decompose

$$S_{\sigma_0} = \frac{U_n^2 \Delta\hat{\sigma}^2}{\hat{d}_{\sigma_0}\, \epsilon_n \exp(T\sigma_0^2 U_n^2/2)} + \frac{U_n^2(\sigma^2 - \sigma_0^2)}{\hat{d}_{\sigma_0}\, \epsilon_n \exp(T\sigma_0^2 U_n^2/2)}. \tag{6.3}$$

We will show that for $\sigma \leqslant \sigma_0 - \tau$ the first term converges uniformly in probability to zero. The approximation error contributes a term that converges deterministically to zero by the bound (4.29) and by the assumption $\epsilon_n U_n^{s+1} \exp(T\sigma_0^2 U_n^2/2) \to \infty$ as $n \to \infty$. The remainder term of the stochastic error in the first summand of (6.3) converges uniformly in probability to zero by Section 5.4 and the linearized stochastic error by the following lemma.

Lemma 6.4. *Let $\delta \in L^\infty(\mathbb{R})$. Let $w_U \in L^\infty([0,1], \mathbb{C})$ be Riemann–integrable and let* *there be a constant $C > 0$ such that $\|w_U\|_\infty \leqslant C$ for all $U \geqslant 1$. Then for all $\kappa, \tau > 0$*

$$\sup_{\mathcal{T}\in\mathcal{H}_s(R,\sigma_0-\tau)} \mathbb{P}_{\mathcal{T}}\left(\left|\frac{1}{\epsilon_n \exp(T\sigma_0^2U^2/2)}\int_0^1 w_U(u)\mathcal{L}_{n,U}(u)\mathrm{d}u\right| > \kappa\right) \to 0,$$

as $U \to \infty$

Proof. It is equivalent to consider for all $\tau' > 0$ the supremum over $\mathcal{H}_s(R, \sqrt{\sigma_0^2 - \tau'})$, such that $\sigma^2 \leqslant \sigma_0^2 - \tau'$ is satisfied for each $\mathcal{T}$. By (4.16) we have

$$\begin{aligned}
&\sup_{\mathcal{T}} \frac{1}{\epsilon_n^2 \exp(T\sigma_0^2U^2)}\mathbb{E}\left[\int_0^1 w_U(u)\mathcal{L}_{n,U}(u)\mathrm{d}u\overline{\int_0^1 w_U(v)\mathcal{L}_{n,U}(v)\mathrm{d}v}\right]\\
&= \sup_{\mathcal{T}} \frac{2\pi U^4 \exp(2T(\lambda-\gamma-\sigma^2/2))}{T^2\exp(T\sigma_0^2U^2)}\int_0^1\int_0^1 f_U(u)\overline{f_U(v)}\\
&\quad \overline{\mathcal{F}^{-1}(\delta(y+\theta)^2)(U(u-v))}e^{T\sigma^2U^2(u^2+v^2)/2}\mathrm{d}u\mathrm{d}v\\
&\leqslant \frac{4\pi U^4\exp(6TR)C^2}{T^2\exp(T\tau' U^2)}\|\mathcal{F}^{-1}\delta^2\|_\infty \to 0,
\end{aligned}$$

as $U \to \infty$, where we used $\|f_U\|_\infty \leqslant \sqrt{2}C\exp(TR)$. □

We have seen that for $\sigma \leqslant \sigma_0 - \tau$ the first term in (6.3) converges to zero uniformly in probability. The second term is $\hat{d}_{\sigma_0}^{-1}$ times a deterministic sequence converging to $-\infty$. We note that $\hat{d}_{\sigma_0}$ is bounded from above and below. Consequently, it holds $\lim_{n\to\infty}\inf_{\mathcal{H}_s(R,\sigma_0-\tau)}\mathbb{P}(S_{\sigma_0} < c) = 1$ for all $c \in \mathbb{R}$. We would like to make a similar statement in the case $\sigma \geqslant \sigma_0 + \tau$. Unfortunately for $\sigma > \sigma_0$ the variance of $\mathcal{L}_{n,U_n}(1)$ does not converge to zero. So it is not possible to find a bound like in Proposition 3.3, which converges to zero. Consequently, the remainder term cannot be controlled by Lemma 4.10. We modify the test statistic in the following way. We choose $\bar{\sigma} > \sigma_{\max}$ and let $\bar{U}_n$ be a cut–off value with $\epsilon_n\bar{U}_n^2\exp(T\bar{\sigma}^2\bar{U}_n^2/2) \to 0$ and $\epsilon_n\bar{U}_n^{s+1}\exp(T\bar{\sigma}^2\bar{U}_n^2/2) \to \infty$ as $n \to \infty$. We further assume that

$$\frac{\bar{U}_n^2\exp(-T\bar{\sigma}^2\bar{U}_n^2/2)}{U_n^2\exp(-T\sigma_0^2U_n^2/2)} \to \infty. \tag{6.4}$$

This can for example be ensured by choosing the cut–off values U_n and $\bar{U}_n$ according to (4.6) with α and $\bar{\alpha} > \alpha$, respectively. Let $\widetilde{\sigma}^2$ be the estimator of σ^2 corresponding to the cut–off value $\bar{U}_n$. We define

$$\begin{aligned}
\widetilde{S}_{\sigma_0} &:= S_{\sigma_0} + \frac{\bar{U}_n^2(\widetilde{\sigma}^2 - \sigma_0^2)}{\epsilon_n\exp(T\bar{\sigma}^2\bar{U}_n^2/2)} &(6.5)\\
&= S_{\sigma_0} + \frac{\bar{U}_n^2(\widetilde{\sigma}^2 - \sigma^2)}{\epsilon_n\exp(T\bar{\sigma}^2\bar{U}_n^2/2)} + \frac{\bar{U}_n^2(\sigma^2 - \sigma_0^2)}{\epsilon_n\exp(T\bar{\sigma}^2\bar{U}_n^2/2)}. &(6.6)
\end{aligned}$$

Under H_0 the statistic S_{σ_0} can be written as in Lemma 5.4. The second term in (6.5) converges uniformly in probability to zero. Thus, under H_0 the modified statistic $\widetilde{S}_{\sigma_0}$ converges uniformly in distribution to a standard normal random variable by Lemma 5.4. For $\sigma \leqslant \sigma_0 - \tau$ the second term in (6.6) converges uniformly in probability to zero and the third term is deterministic sequence converging to $-\infty$. As for S_{σ_0} it holds $\lim_{n\to\infty} \inf_{\mathcal{H}_s(R,\sigma_0-\tau)} \mathbb{P}(\widetilde{S}_{\sigma_0} < c) = 1$ for all $c \in \mathbb{R}$. But now we are also able to make a similar statement for $\sigma \geqslant \sigma_0 + \tau$. Since we bounded the estimators S_{σ_0} cannot diverge faster than $\epsilon_n^{-1} U_n^2 \exp(-T\sigma_0^2 U_n^2/2)$. For $\sigma \in [0, \sigma_{\max}]$ the second term in (6.6) converges uniformly in probability to zero. Owing to (6.4) the third term in (6.6) tends to infinity faster than the bound of S_{σ_0}. It holds $\lim_{n\to\infty} \inf_{\mathcal{T}} \mathbb{P}(\widetilde{S}_{\sigma_0} > c) = 1$ for all $c \in \mathbb{R}$, where the infimum is over all $\mathcal{T} \in \mathcal{H}_s(R, \sigma_{\max})$ with $\sigma \geqslant \sigma_0 + \tau$.

Let q_η denote the $(1-\eta)$–quantile of the standard normal distribution. For $\sigma_0 \in (0, \sigma_{\max})$ we define the tests

$$\varphi_{\sigma_0} := \begin{cases} 0, & \text{if } |\widetilde{S}_{\sigma_0}| \leqslant q_{\eta/2} \\ 1, & \text{if } |\widetilde{S}_{\sigma_0}| > q_{\eta/2} \end{cases}, \qquad \varphi_{\sigma_{\max}} := \begin{cases} 0, & \text{if } S_{\sigma_{\max}} \geqslant q_{1-\eta} \\ 1, & \text{if } S_{\sigma_{\max}} < q_{1-\eta} \end{cases}.$$

For $\sigma_0 = 0$ we would like to apply Theorem 4.1. To this end, we choose the cut–off value U_n such that $\epsilon_n U_n^{5/2} \to 0$ and $\epsilon_n U_n^{(2s+5)/2} \to \infty$. Let $\hat{\sigma}^2$ be the estimator corresponding to this cut–off value U_n, where $\hat{\sigma}^2 \in [0, \sigma_{\max}^2]$ is ensured. We assume that δ is positive on $[-TR, TR]$ and define

$$\begin{aligned} S_0 &:= \frac{U_n^{1/2} \hat{\sigma}^2}{\hat{d}_0\, \epsilon_n} + \frac{\bar{U}_n^2 \widetilde{\sigma}^2}{\epsilon_n \exp(T\sigma_{\max}^2 \bar{U}_n^2/2)}, \\ \hat{d}_0 &:= \frac{2\sqrt{\pi}\delta(T\hat{\gamma})}{\exp(T(\hat{\gamma} - \hat{\lambda}))T} \left(\int_0^1 u^4 w_\sigma^1(u)^2 \mathrm{d}u \right)^{1/2}. \end{aligned}$$

Under $H_0 : \sigma^2 = 0$ the test statistic S_0 converges uniformly in distribution to a standard normal random variable by a similar argument as for $\widetilde{S}_{\sigma_0}$. We observe that the first term of S_0 is nonnegative. Under $H_1 : \sigma \geqslant \tau$ the second term of S_0 may be decomposed as in (6.6) into a part that converges uniformly in distribution to zero and into a deterministic sequence converging to infinity. It holds $\lim_{n\to\infty} \inf_{\mathcal{T}} \mathbb{P}(S_0 > c) = 1$ for all $c \in \mathbb{R}$, where the infimum is over all $\mathcal{T} \in \mathcal{H}_s(R, \sigma_{\max})$ with $\sigma \geqslant \tau$. We define the test

$$\varphi_0 := \begin{cases} 0, & \text{if } S_0 \leqslant q_\eta \\ 1, & \text{if } S_0 > q_\eta \end{cases}.$$

Since we have a test on $\sigma = \sigma_0$ for each $\sigma_0 \in [0, \sigma_{\max}]$, we may use this family of tests to define a confidence set for σ. We define $M_n := \{\sigma | \varphi_\sigma = 0\}$ and obtain $\lim_{n\to\infty} \inf_{\mathcal{T}} \mathbb{P}_{\mathcal{T}}(\sigma \in M_n) = 1 - \eta$, where the infimum is over all $\mathcal{T} \in \mathcal{H}_s(R, \sigma_{\max})$ with volatility σ. The set M_n is not necessarily an interval. For $\sigma_0 \in (0, \sigma_{\max})$ the cut–off value U_n may be chosen as a continuous function of σ_0 by (4.6). The estimators $\hat{\sigma}^2$, $\hat{\gamma}$ and $\hat{\lambda}$ depend continuously on the cut–off value U_n, which can be seen by substituting $v = u/U$ in (2.12), (2.13) and (2.14) and applying the continuity theorem on parameter dependent integrals. Thus, $\widetilde{S} : (0, \sigma_{\max}) \to \mathbb{R}$, $\sigma \mapsto \widetilde{S}_\sigma$ is continuous and $M_n \cap (0, \sigma_{\max})$ may be written as the preimage $\widetilde{S}^{-1}([-q_{\eta/2}, q_{\eta/2}])$ of the continuous function $\widetilde{S}$.

7 Simulations and empirical results

In the last chapter, we derived asymptotic confidence intervals for the volatility, the drift, the intensity and pointwise for the jump density. This chapter is concerned with the application of the results in simulations and to option data of the German DAX index. It is based on the results that will appear in Söhl and Trabs (2012b), where in addition to finite intensity Lévy processes also self–decomposable Lévy processes are treated. More precisely, the paper is complemented by confidence intervals for the method by Trabs (2012) and their application in simulations and to option data. We slightly adjust the theoretical concepts of the last chapter and thus enhance the performance of the method in practice. The main difference concerns the construction of the confidence sets. Applying the confidence intervals in simulations with sample sizes as in available data shows that they are to conservative. In the proof of the asymptotic normality, we have seen that the asymptotic variances are completely determined by the linearized stochastic errors and that the remainder terms and the approximation errors are asymptotically negligible. Replacing the asymptotic variances of the linearized stochastic errors by the finite sample variances certainly gives a more precise measure of how the estimators deviate from the true value. So we base our confidence intervals on the finite sample variances of the linearized stochastic errors. However, this still leads to asymptotic and not to finite sample confidence intervals since we continue to take advantage of the remainder terms and the approximation errors being asymptotically negligible. The so constructed intervals perform well in terms of size and coverage probabilities as we will demonstrate by simulations from the model by Merton (1976).

Our estimators for (σ^2, γ, ν) are constructed essentially as described in Section 2.2, but some modifications are introduced which improve their numerical performance. As shown in simulations these improvements reduce the mean squared error of the estimators significantly. In contrast to the method by Cont and Tankov (2004b) the spectral calibration method is a straightforward algorithm, where no minimization problem has to be solved. Therefore, the method is quite fast owing to the Fast Fourier transform (FFT). In this chapter, we concentrate on the application of the method to realistic sample sizes. In a related framework of a jump–diffusion Libor model, Belomestny and Schoenmakers (2011) study the application of the spectral calibration method to finite sample data sets.

We use data of vanilla options on the German DAX index. Considering options with different maturities, the model achieves good calibration results in the sense that the residuals between the given data and the calibrated model are small. Since the Blumenthal–Getoor index equals zero in our model, the calibration based on option data behaves quite differently from the case of high–frequency observations under the historical measure, where Aït-Sahalia and Jacod (2009) find evidence that the Blumenthal–

Getoor index is larger than one. Applying the calibration to a sequence of trading days, we obtain the evolution of the model parameters in time. The estimators seem to be stable with respect to the spot time.

This chapter is organized as follows: In Section 7.1 we specify the estimation method used in the simulations and for the option data of the German DAX index. In Section 7.3 the confidence intervals are derived and their performance is assessed in simulations. We apply the method to the data and discuss our results in Section 7.4. The more technical part of determining the finite sample variances is deferred to Section 7.5.

7.1 The estimation method in applications

A typical parametric submodel is given by Example 7.1. We will use it to study the performance of estimation method in simulations.

Example 7.1 (Merton model)**.** Merton (1976) introduced the first exponential Lévy model. Therein, the jumps are normally distributed with intensity $\lambda > 0$:

$$\nu(x) = \frac{\lambda}{\sqrt{2\pi}v} \exp\left(-\frac{(x-\eta)^2}{2v^2}\right), \quad x \in \mathbb{R}.$$

A realistic choice of the parameters is $\eta = -0.1$, $v = 0.2$ and $\lambda = 5$. Together with the volatility $\sigma = 0.1$ this determines the drift to be $\gamma = 0.379$ using the martingale condition (2.8).

We will use the estimators $\hat{\sigma}^2$, $\hat{\gamma}$ and $\hat{\lambda}$ as defined in (2.12)–(2.14). In this chapter, we will use the estimator $\hat{\nu}$ obtained by plugging in $\hat{\sigma}^2$, $\hat{\gamma}$ and $\hat{\lambda}$ in the estimation formula (2.23) for ν of the slightly modified estimation method. This has the advantage that ν can be estimated directly without estimating first μ by (2.16) and then multiplying by e^{-x}. This direct estimation turns out to be more stable in simulations with small sample sizes.

Note that correction steps are necessary to satisfy non–negativity of the jump density and the martingale condition (2.8). If the latter one would be violated, the right–hand side of the pricing formula (2.9) could have a singularity at zero and thus we could not apply the inverse Fourier transform to obtain an option function $\mathcal{O}$ from the calibration.

A critical question is the choice of the regularization parameter U. As a benchmark, we use in simulations an oracle cut–off value, that is U minimizes the discrepancy between the estimators and the true values of σ^2, γ and ν measured in an L^2–loss. To calibrate real data, we employ the simple least squares approach

$$U^* := \operatorname{arginf}_U RSS(U) \quad \text{with the residual sum of squares}$$
$$RSS(U) := \sum_{j=1}^n |\hat{\mathcal{O}}_U(x_j) - O_j|^2, \tag{7.1}$$

where $\hat{\mathcal{O}}_U$ is the option function corresponding to the characteristic triplet estimated by means of the cut–off value U. We determine $\hat{\mathcal{O}}_U$ by the pricing formula (2.9) and Lévy-Khintchine representation (2.7), in which we plug in the estimators obtained by using the cut–off value U. The estimated option function $\hat{\mathcal{O}}_U$ can be computed efficiently for each U so that the numerical effort of finding U^* is mainly determined by the minimization algorithm used to solve (7.1). From theoretical consideration a penalty term, as used by Belomestny and Reiß (2006b), is necessary to avoid an over–fitting, that is not to choose U too large. Nevertheless, our practical experience with this method shows that the above mentioned correction steps, which are not included in the theory, lead to an auto–penalization: Using large cut–off values, the stochastic error in the estimators becomes large. This leads to high fluctuations of the nonparametric part and thus the correction has an increasing effect. Hence, the difference between $\widetilde{\mathcal{O}}$ and $\hat{\mathcal{O}}_U$ becomes larger if U is too high and thus the residual sum of squares increases, too. In particular, imposing the jump density to be nonnegative implies a shape constraint on the state price density which is basically the second derivative of the option function. Therefore, the least squares choice of the tuning parameter works well at least for small noise levels.

The approach to minimize the calibration error was also applied by Belomestny and Schoenmakers (2011). Alternative data–driven choices of the cut–off value U are the "quasi–optimality" approach which was studied by Bauer and Reiß (2008) and which was applied by Belomestny (2011) or the use of a preestimator as proposed by Trabs (2012). However, we will consider only the least squares approach, which performs well in our application.

It remains the choice of the weight functions w_σ^U, w_γ^U and w_λ^U. The estimators in (2.12),(2.13) and (2.14) can be understood as weighted L^2–projections of ψ_n onto the space of quadratic polynomials. In this sense, the estimators arise naturally as a solution of a weighted least squares problem. However, the optimal weight depends not only on the unknown heteroscedasticity in the frequency domain but also on the unknown function $\mathcal{F}\mu$, so we do not pursue this approach here. Instead we construct the weight functions w_σ^U, w_γ^U and w_λ^U directly as Belomestny and Reiß (2006b), but propose weight functions different to theirs (2.24)–(2.26). The idea is that the noise is particularly large in the high frequencies and thus it is desirable to assign less weight to the high frequencies. A smooth transition of the weight functions to zero at the cut–off value improves the numerical results significantly. Therefore, we would like the weight function and its first two derivatives to be zero at the cut–off value. With the side conditions on the weight functions this leads to the following polynomials:

$$\begin{aligned} w_\sigma^U(u) &:= \frac{c_\sigma}{U^3}\Big((2s+1)\left(\frac{u}{U}\right)^{2s} - 4(2s+3)\left(\frac{u}{U}\right)^{2s+2} + 6(2s+5)\left(\frac{u}{U}\right)^{2s+4} \\ &\qquad - 4(2s+7)\left(\frac{u}{U}\right)^{2s+6} + (2s+9)\left(\frac{u}{U}\right)^{2s+8}\Big)\mathbf{1}_{[-U,U]}(u), \\ w_\gamma^U(u) &:= \frac{c_\gamma}{U^2}\left(\left(\frac{u}{U}\right)^{2s+1} - 3\left(\frac{u}{U}\right)^{2s+3} + 3\left(\frac{u}{U}\right)^{2s+5} - \left(\frac{u}{U}\right)^{2s+7}\right)\mathbf{1}_{[-U,U]}(u), \end{aligned}$$

$$w_\lambda^U(u) := \frac{c_\lambda}{U}\Big((2s+3)\left(\frac{u}{U}\right)^{2s} - 4(2s+5)\left(\frac{u}{U}\right)^{2s+2} + 6(2s+7)\left(\frac{u}{U}\right)^{2s+4} - 4(2s+9)\left(\frac{u}{U}\right)^{2s+6} + (2s+11)\left(\frac{u}{U}\right)^{2s+8}\Big)\mathbf{1}_{[-U,U]}(u).$$

The constants $c_\sigma, c_\gamma, c_\lambda \in \mathbb{R}$ are determined by the normalization conditions

$$\int_{-U}^{U} \frac{-u^2}{2} w_\sigma^U(u)\,\mathrm{d}u = 1, \quad \int_{-U}^{U} u w_\gamma^U(u)\,\mathrm{d}u = 1 \quad \text{and} \quad \int_{-U}^{U} w_\lambda^U(u)\,\mathrm{d}u = 1.$$

The parameter s reflects the a priori knowledge about the smoothness of ν and can be chosen equal to two. The gain of the new weight functions is discussed in Section 7.2.

To estimate directly the jump density ν and not only the exponential scaled version μ, we use $\psi(u+i)$ instead of $\psi(u)$ as discussed above. Therefore, we define the estimator

$$\hat{\nu}(x) := \mathcal{F}^{-1}\Big[\Big(\psi_n(u+i) + \frac{\hat{\sigma}^2}{2}u^2 - i\hat{\gamma}u + \hat{\lambda}\Big) w_\nu^U(u)\Big](x), \tag{7.2}$$

where $\psi_n(u+i)$ is the empirical version of $\psi(u+i)$ and w_ν^U is a flat top kernel with support $[-U,U]$:

$$w_\nu^U(u) := F\left(\frac{u}{U}\right) \quad \text{with} \quad F(u) := \begin{cases} 1, & |u| \le 0.05, \\ \exp\left(\frac{-\exp(-(|u|-0.05)^{-2})}{(|u|-1)^2}\right), & 0.05 < |u| < 1, \\ 0, & |u| \ge 1. \end{cases} \tag{7.3}$$

To evaluate the integrals in (2.12) to (2.14), it suffices to apply the trapezoidal rule. The inverse Fourier transformation in (7.2) can be efficiently computed using the FFT–algorithm. Therefore, depending on the interpolation method which is applied to obtain $\mathcal{O}_n$, the whole estimation procedure is very fast. Finally, we note that the cut–off value can be chosen differently for each quantity $\sigma^2, \gamma, \lambda$ and ν.

7.2 Simulations

Let us first describe the setting of all of our simulations. In view of the higher concentration of European options at the money, the design points $\{x_1, \dots, x_n\}$ are chosen to be the $k/(n+1)$–quantiles, $k = 1, \dots, n$, of a normal distribution with mean zero and variance 1/2. The observations O_j are computed from the characteristic function φ_T using the fast Fourier transform. The additive noise consists of independent, normal and centered random variables with variance $\delta_j^2 = |\tau\mathcal{O}(x_j)|^2$ for some relative noise level $\tau > 0$. We use the notation $\|\delta\|_{l^\infty} := \sup_j \delta_j$. By choosing the sample size n and the deviation parameter τ, we determine the noise level of the observations. According to the existing theoretical results by Belomestny and Reiß (2006a) as well as by Trabs (2012),

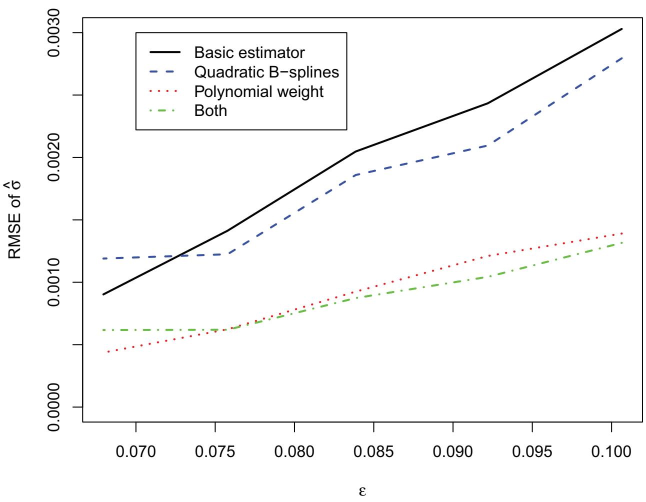

Figure 7.1: RMSE of $\hat{\sigma}$ for different noise levels with 500 Monte–Carlo iterations in each case. Usage of the linear and quadratic spline interpolation as well as usage of the weight functions by Belomestny and Reiß (2006a) and the polynomial weight functions.

it is well measured by the quantity

$$\varepsilon := \Delta^{3/2} + \Delta^{1/2}\|\delta\|_{l^\infty} \quad \text{with} \quad \Delta := \max_{j=2,\dots,n}(x_j - x_{j-1}),$$

which takes the interpolation error and the stochastic error into account. The interest rate and time to maturity are set to $r = 0.06$ and $T = 0.25$, respectively.

Using the Merton model with the parameters of Example 7.1, we investigate the practical influence of two aspects of the procedure. The interpolation of the data (x_j, O_j) with linear B–splines is compared to the use of quadratic B–splines. The latter preprocessing is an additional smoothing of the data, which achieves significant gains for higher noise levels. The other point of interest is the choice of the weight functions. Since it is known from the theory that the noise affects mainly the high frequencies, the polynomial weight functions greatly reduce the variance of the estimator. These improvements are illustrated in Figure 7.1: In the case of $\hat{\sigma}$ we approximate the root mean squared error (RMSE) $\sqrt{\mathbb{E}[|\hat{\sigma} - \sigma|^2]}$ using 500 Monte–Carlo iterations with and without quadratic splines and polynomial weight functions, respectively. This is done for different noise levels, whereby τ decreases from 0.03 to 0.015 and n increases from 50 to 400, simultaneously. Further simulation results, in particular for estimating the jump density, can be found in Section 7.3.

7.3 Confidence intervals

While we have constructed confidence intervals based on the asymptotic variance in Section 6.1, we want to construct in this section confidence intervals based on the finite sample variance of the linearized stochastic errors. The construction is in the spirit of the asymptotic analysis of Chapter 4 in the sense that the remainder terms and the approximation errors are not taken into account. But given this, it is an independent, rather direct approach to the construction of confidence sets, which transfers easily to similar methods. Presented in a less rigorous manner, it is a practical method to obtain confidence sets.

Let us consider $\hat{\sigma}^2$ first. All other parameters can be treated similarly. By the undersmoothing approach to the construction of confidence intervals, we neglect the approximation error and consider the stochastic error only, compare (2.29),

$$\hat{\sigma}^2 - \sigma^2 \approx \int_{-U}^{U} \mathrm{Re}(\Delta\psi_n(u)) w_\sigma^U(u)\,\mathrm{d}u \tag{7.4}$$

with $\Delta\psi_n := \psi_n - \psi$. The term $\Delta\psi_n$ is a logarithm, which we approximate by

$$\Delta\psi_n(u) = \frac{1}{T}\log\left(\frac{1 + iu(1+iu)\,\mathcal{F}\,\mathcal{O}_n(u)}{1 + iu(1+iu)\,\mathcal{F}\,\mathcal{O}(u)}\right) \approx \frac{iu(1+iu)}{T\varphi_T(u-i)}(\mathcal{F}\,\mathcal{O}_n - \mathcal{F}\,\mathcal{O})(u) \tag{7.5}$$

using $\log(1+x) \approx x$ for small x. Apply the approximation (7.5) to the right–hand side of (7.4) yields the linearized stochastic error.

To analyze the deviation $\mathcal{F}\,\mathcal{O}_n - \mathcal{F}\,\mathcal{O}$ in the linearized stochastic error, we assume that the noise levels of the observations (2.6) are given by the values $\delta_j = \delta(x_j)$, $j = 1, \ldots, n$, of some function $\delta : \mathbb{R} \to \mathbb{R}_+$. The observation points are assumed to be the quantiles $x_j = H^{-1}(j/(n+1))$, $j = 1, \ldots, n$, of a distribution with c.d.f. $H : \mathbb{R} \to [0, 1]$ and p.d.f. h. For the definition of the confidence intervals we need the generalized noise level

$$\varrho(x) = \delta(x)/\sqrt{h(x)}, \tag{7.6}$$

which incorporates the noise of the observations as well as their distribution. Instead of assuming that the observation points are given by the quantiles of h one may also assume that the observation points are sampled randomly from the density h. On these conditions Brown and Low (1996) showed the asymptotic equivalence in the sense of Le Cam of the nonparametric regression model (2.6) and the white noise model $\mathrm{d}Z_n(x) = \mathcal{O}(x)\,\mathrm{d}x + n^{-1/2}\varrho(x)\,\mathrm{d}W(x)$ with a two–sided Brownian motion W. This asymptotic equivalence has been discussed where the Gaussian white noise model (2.30) is introduced in Chapter 2. Z_n is an empirical version of the antiderivative of $\mathcal{O}$. In that sense we define

$$\mathcal{F}\,\mathcal{O}_n(u) := \mathcal{F}[\,\mathrm{d}Z_n](u) = \mathcal{F}\,\mathcal{O}(u) + n^{-1/2}\int_{\mathbb{R}} e^{iux}\varrho(x)\,\mathrm{d}W(x).$$

Combining (7.5) with this asymptotic equivalence, we can approximate

$$\Delta\psi_n(u) \approx \frac{1}{\sqrt{n}} \frac{iu(1+iu)}{T\varphi_T(u-i)} \int_{\mathbb{R}} e^{iux} \varrho(x) \,\mathrm{d}W(x) =: \frac{1}{\sqrt{n}} \mathcal{L}(u).$$

Defining $f_{\sigma,U}(u) := w_\sigma^U(u) iu(1+iu)/(T\varphi_T(u-i))$, we obtain with the above considerations and by changing the order of integration

$$\hat{\sigma}^2 - \sigma^2 \approx \frac{1}{\sqrt{n}} \int_{-U}^{U} \mathrm{Re}(\mathcal{L}(u)) w_\sigma^U(u) \,\mathrm{d}u = \frac{2\pi}{\sqrt{n}} \int_{\mathbb{R}} \mathrm{Re}\left(\mathcal{F}^{-1} f_{\sigma,U}(-x)\right) \varrho(x) \,\mathrm{d}W(x).$$

For σ^2 we calculate the finite sample variance $s_{\sigma^2}^2$ of the linearized stochastic errors using the Itô isometry

$$\begin{aligned} s_{\sigma^2}^2 &= \frac{1}{n} \mathbb{E}\left[\left(\int_{-U}^{U} \mathrm{Re}(\mathcal{L}(u)) w_\sigma^U(u) \,\mathrm{d}u\right)^2\right] \\ &= \frac{4\pi^2}{n} \mathbb{E}\left[\left(\int_{\mathbb{R}} \mathrm{Re}\left(\mathcal{F}^{-1} f_{\sigma,U}(-x)\right) \varrho(x) \,\mathrm{d}W(x)\right)^2\right] \\ &= \frac{4\pi^2}{n} \int_{\mathbb{R}} \left(\mathrm{Re}\left(\mathcal{F}^{-1} f_{\sigma,U}(-x)\right) \varrho(x)\right)^2 \mathrm{d}x. \end{aligned} \tag{7.7}$$

Similar results for $\hat{\gamma}, \hat{\lambda}$ and $\hat{\nu}(x_0), x_0 \in \mathbb{R}$, are derived in Section 7.5. The corresponding finite sample variances s_γ^2, s_λ^2 and $s_{\nu(x_0)}^2$ are given by (7.9), (7.10) and (7.11), respectively. In contrast to the asymptotic normality results in Chapter 4, we do not have to distinguish between the cases $x_0 = 0$ and $x_0 \neq 0$ in the finite sample analysis. An advantage of this way of deriving the variance and thus of constructing confidence sets is that it can easily be transferred to similar methods. For the calibration of self–decomposable Lévy processes this is done in Söhl and Trabs (2012b).

To construct confidence intervals for $\vartheta \in \{\sigma^2, \gamma, \lambda, \nu(x_0)\}$, we need an estimate $\hat{s}_\vartheta$ of the standard deviation. To this end, the function $f_{q,U}$ has to be replaced by its empirical version. Since the only unknown quantity involved is φ_T, it suffices to plug in an estimator for the characteristic function. Either one uses the Lévy-Khintchine representation (2.7) replacing the true characteristic triplet by their estimators or φ_T is estimated by the empirical version ψ_n in (2.11) and (2.19), respectively. We will follow the second approach since the estimate is independent of the cut–off value U and thus may lead to more stable results. To compute the noise function ϱ, the density h of the distribution of the strikes is necessary while in practice there is only a discrete set of strike prices. The density h can be estimated from the observation points $(x_j)_{j=1,\dots,n}$ using some standard density estimation method. We will apply a triangular kernel estimator, where the bandwidth is chosen by Silverman's rule of thumb. The $(1-\eta)$–confidence intervals for a level $\eta > 0$ are then given by

$$I_\vartheta := [\hat{\vartheta} - \hat{s}_\vartheta q_{\eta/2}, \hat{\vartheta} + \hat{s}_\vartheta q_{\eta/2}], \tag{7.8}$$

where q_η denotes the $(1-\eta)$–quantile of the standard normal distribution. Naturally,

	σ^2	γ	λ	$\nu(x_0)$
U	54	50	46	26
$\eta = 0.5$	53%	48%	43%	48%
$\eta = 0.05$	94%	93%	81%	91%

Table 7.1: Approximate coverage probabilities of $(1-\eta)$–confidence intervals from a Monte Carlo simulation with 1000 iterations and fixed cut–off values U. The confidence interval of ν is evaluated at $x_0 = -0.2$.

both the estimator $\hat{\vartheta}$ and the size of the confidence set, determined by $\hat{s}_\vartheta$, depend the choice of the cut–off value U. In particular, it reflects the bias–variance trade–off of the estimation problem: Small values of U lead to a strong smoothing such that the interval (7.8) will be small but there might be a significant bias. Using larger U, the confidence intervals become wider but the deterministic error reduces. Therefore, only by undersmoothing the interval (7.8) has asymptotically the level η.

In practice we are rather interested in the parameter σ than in its square. Applying the delta method, the finite sample variance of the estimator $\hat{\sigma} := \sqrt{\hat{\sigma}^2}$ is given by $s_\sigma^2 = \frac{1}{4}s_{\sigma^2}^2\sigma^{-2}$ and thus its empirical version is $\hat{s}_\sigma^2 = \frac{1}{4}\hat{s}_{\sigma^2}^2(\hat{\sigma}^2)^{-1}$. This allows to construct confidence intervals for $\hat{\sigma}$, too.

We examine the performance of the confidence intervals by simulations from the Merton model with parameters as in Example 7.1. As in Section 7.2, the interest rate is chosen as $r = 0.06$ and the time to maturity as $T = 0.25$. We simulate $n = 100$ strike prices and take the relative noise level to be $\tau = 0.01$. To coincide with the theory, we interpolate the corresponding European call prices linearly. In the real data application in Section 7.4 we will take advantage of the B-spline interpolation.

We asses the performance of the confidence intervals (7.8) with levels $\eta = 0.5$ and $\eta = 0.05$ in a Monte Carlo simulation with 1000 iterations in each model. The cut–off values are fixed for any quantity and larger than the oracle choice of U. This ensures that the bias is indeed negligible. As a rule of thumb the cut–off values for the confidence sets can be chosen to be 4/3 of the oracle cut–off value. We approximate the coverage probabilities of the confidence sets by the percentage of confidence intervals which contain the true value in a finite number of cases. Table 7.1 gives the chosen cut–off values and the approximate coverage probabilities. Further simulations show that for sufficiently small levels, for instance $\eta = 0.05$, the confidence intervals have a reasonable size for a wide range of cut–off values. However, the parameter λ falls a bit out of the general picture.

Based on simulations in the Merton model, Figure 7.2 illustrates the true jump density, its estimator with oracle choice of the cut–off values and the corresponding pointwise 95% confidence intervals. Almost everywhere the true function is contained in the confidence intervals. Moreover, another 100 estimators from further Monte Carlo iterations are plotted. The graph shows that the confidence intervals describe well the deviation of the estimated jump densities. The negative bias around zero might come from the smoothing which naturally tends to smooth out peaks, cf. (Härdle, 1990, Chap. 5.3).

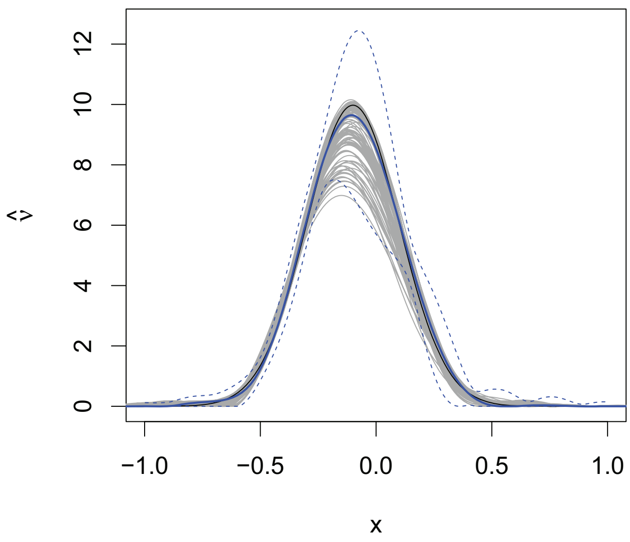

Figure 7.2: True *(solid)* and estimated *(bold)* jump density with pointwise 95% confidence intervals *(dashed)*, using the oracle cut–off value $U = 19$. Additional 100 estimated jump densities *(grey)* from a Monte Carlo simulation of the Merton model.

7.4 Empirical study

We apply the calibration methods to a data set from the Deutsche Börse database Eurex[1]. It consists of settlement prices of European put and call options on the German DAX index from May 2008. Therefore, the prices are observed before the latest financial crises and thus the market activity is stable. The interest rate r is chosen for each maturity separately according to the put–call parity at the respective strike prices. The expiry months of the options are between July and December, 2008, and thus the time to maturity T, which we measure in years, reaches from two to seven months. The number of our observations n is given in Figure 7.3 and lays around 50 to 100 different strikes for each maturity and trading day.

To apply the confidence intervals (7.8) of Section 7.3, we need the noise function ϱ from (7.6). By a rule of thumb we assume δ to be 1% of the observed prices $\mathcal{O}(x_j)$ (cf. Cont and Tankov, 2004a, p. 439). All other unknown quantities are estimated as discussed above.

Let us first focus on one (arbitrarily chosen) day. Hence, we calibrate the option prices of May 29, 2008, with all four different maturities. The results are summarized

[1]provided by the SFB 649 "Economic Risk"

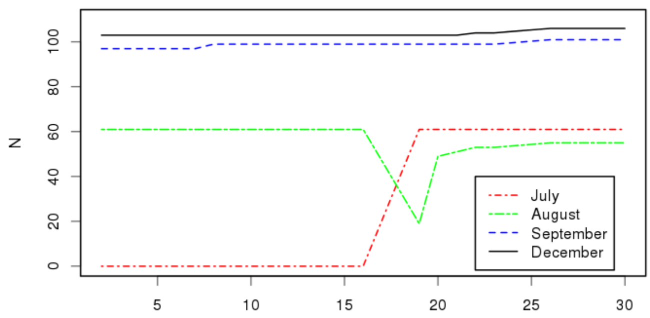

Figure 7.3: Number of observed prices of put and call options during May, 2008.

in Table 7.2 and Figure 7.4. Using the complete estimation of the model, we generate the corresponding option functions $\hat{\mathcal{O}}$. They are graphically compared to the given data points and we calculate the residual sum of squares $RSS = RSS(U^*)$ as defined in (7.1). For all maturities the method yields good fits to the data.

In all maturities the jump density has more weight on the negative half line and thus there are more negative than positive jumps priced into the options. This coincides with the empirical findings in the literature (see eg, Cont and Tankov, 2004a). Due to the positivity correction, the jump densities might look unsmooth, where they are close to zero. This problem might be circumvented by adding smoothness constraints. However, the construction of confidence intervals would then be much more difficult. Hence, this topic is left open for further research.

In view of the parametric calibration of their CGMY model, Carr et al. (2002) suggested that risk–neutral processes of assets should be modeled by pure jump processes with finite variation. Now, the nonparametric approach shows that the model is able to reproduce the option data. Note that the Blumenthal–Getoor index equals zero, which is in contrast to the investigation of high-frequency historical data, where Aït-Sahalia and Jacod (2009) estimated a jump activity index larger than one.

Next we investigate the calibration across different trading days. By considering more than one day we investigate the stability of the estimation procedure. Moreover, calibrating the model across the trading days in May, 2008, shows the development of the model along the time line and with small changes in the maturities. To profit from the higher observation numbers, we apply the calibration procedure to the options with maturity in September and December.

The estimations of the parameters are displayed in Figure 7.5. Note that we do not smooth over time. Furthermore, the 95% confidence intervals for the December options are shown. The estimated volatility $\hat{\sigma}$ fluctuates around 0.1 and 0.12. The confidence sets imply that there is no significant difference between the two maturities. Both $\hat{\gamma}$ and $\hat{\lambda}$ decrease for higher durations: On the one hand the curves of December lie significantly

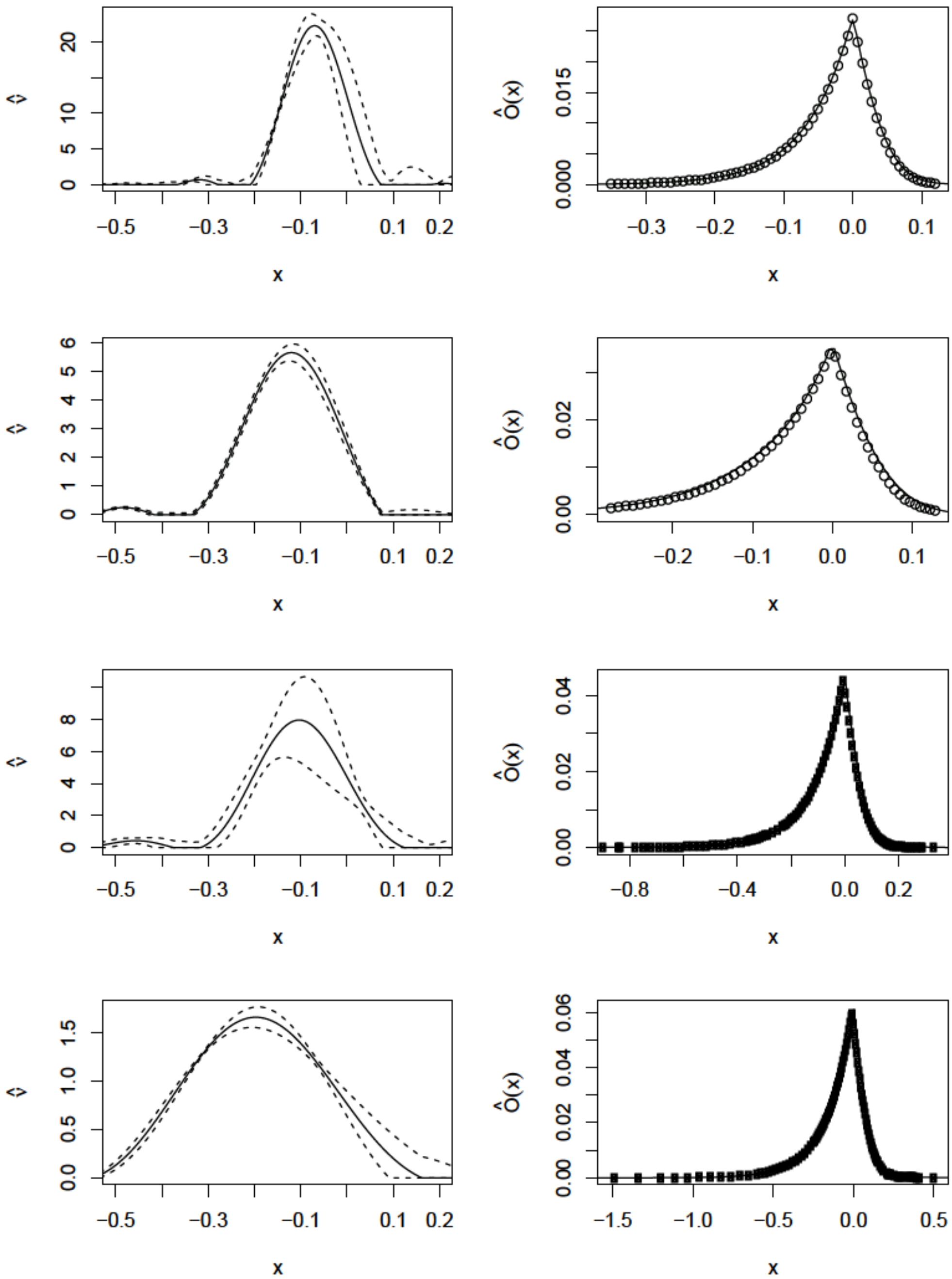

Figure 7.4: Estimated jump densities *(left)* with pointwise 95% confidence intervals as well as calibrated option functions *(right, solid)* and given data from May 29, 2008 *(right, points)*. The time to maturity increases from $T = 0.136$ *(top)* to $T = 0.564$ *(bottom)*.

n	61		55		101		106	
T	0.136		0.233		0.311		0.564	
$\hat{\sigma}$	0.110	(0.0021)	0.123	(0.0009)	0.107	(0.0030)	0.124	(0.0013)
$\hat{\gamma}$	0.221	(0.0049)	0.142	(0.0015)	0.174	(0.0050)	0.105	(0.0011)
$\hat{\lambda}$	3.392	(0.2015)	1.290	(0.0176)	1.823	(0.1261)	0.637	(0.0181)
$\sqrt{RSS}$	0.003		0.008		0.005		0.008	

Table 7.2: Estimated parameters ϑ and estimated standard deviation $\hat{s}_\vartheta$ (in brackets) for $\vartheta \in \{\sigma, \gamma, \lambda\}$ and residual sum of squares using option prices from May 29, 2008, with n observed strikes for each maturity T.

below the ones of September, on the other hand the graphs have a slight positive trend with respect to the time axis, which means with smaller time to maturity. Keeping in mind that the implied volatility in the Black–Scholes model typically decreases for longer time to maturity, this lower market activity is reproduced by smaller jump activities in our calibration while the volatility is relatively stable.

Figure 7.6 displays the estimated jump densities. All Lévy measures have a similar shape, which is in line with real data calibration of Belomestny and Reiß (2006b). In contrast to Cont and Tankov (2004b) the densities are unimodal or have only minor additional modes in the tails, which may be artifacts of the spectral calibration method. The tails of $\hat{\nu}$ do not differ significantly, while the different heights reflect the development of the jump activities $\hat{\lambda}$. There is an obvious trend to small negative jumps in all data sets, which is in line with the stylized facts of option pricing models. The calibration is stable for consecutive market days.

Our empirical investigation shows that the model can be calibrated well to European option prices. Using the derived confidence intervals, we can observe significant changes of the model over time. While the volatility has no systematic trend, the jump activities decrease for longer maturities and thus the jump densities become flatter.

7.5 Finite sample variances of γ, λ and ν

The confidence intervals for γ and λ are based on the finite sample variances of the corresponding linearized stochastic errors. With $f_{\gamma,U}(u) := w_\gamma^U(u) iu(1+iu)/(T\varphi_T(u-i))$ and $f_{\lambda,U}(u) := w_\lambda^U(u) iu(1+iu)/(T\varphi_T(u-i))$ we obtain by definitions (2.13) and (2.14) and the same arguments as in Section 7.3

$$\begin{aligned}\Delta\hat{\gamma} := \hat{\gamma} - \gamma &\approx -\Delta\hat{\sigma}^2 + \int_{\mathbb{R}} \operatorname{Im}\big(\Delta\psi_n(u)\big) w_\gamma^U(u)\,\mathrm{d}u \\ &\approx \frac{2\pi}{\sqrt{n}} \int_{\mathbb{R}} \Big(-\operatorname{Re}\big(\mathcal{F}^{-1} f_{\sigma,U}(-x)\big) + \operatorname{Im}\big(\mathcal{F}^{-1} f_{\gamma,U}(-x)\big)\Big) \varrho(x)\,\mathrm{d}W(x),\end{aligned}$$

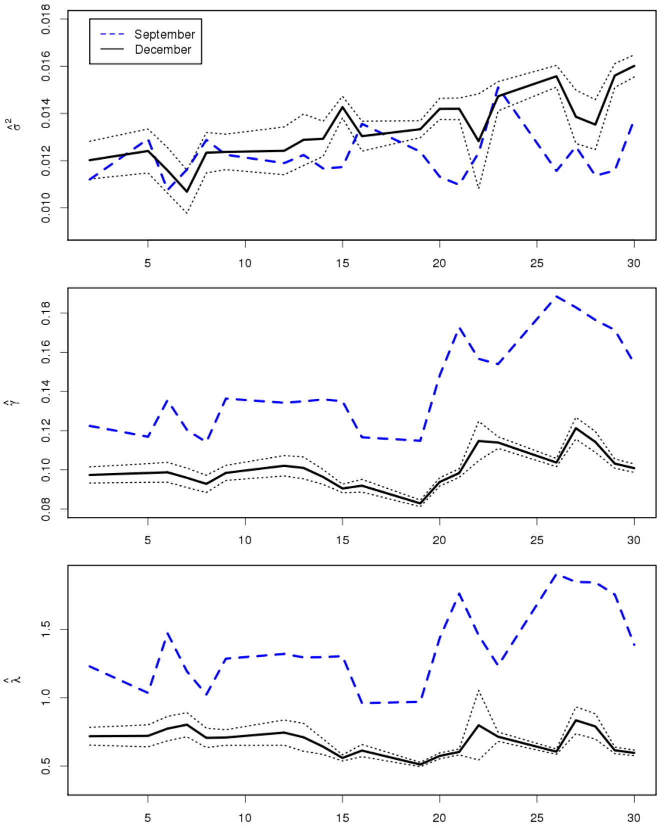

Figure 7.5: At each market day in May, 2008, estimated σ^2 *(top)*, γ *(center)* and λ *(bottom)* from options with maturities in September *(dashed)* and December *(solid)* and confidence intervals *(dotted)* for the latter ones.

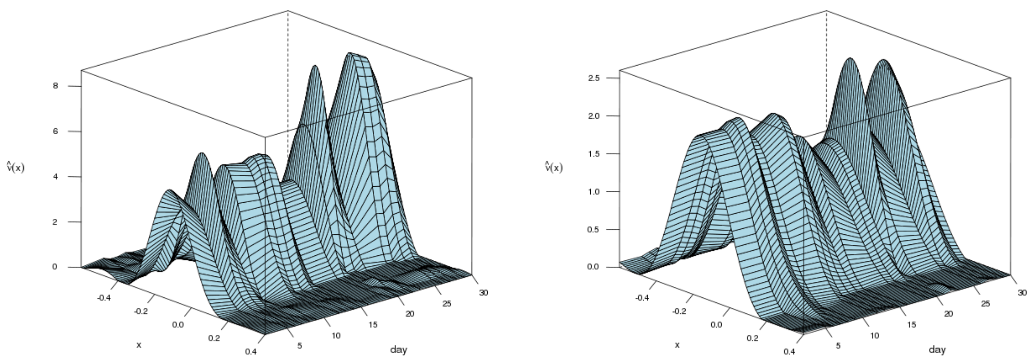

Figure 7.6: Estimation of ν for maturity in September *(left)* and December (*right*).

$$
\begin{aligned}
\Delta\hat{\lambda} := \hat{\lambda} - \lambda &\approx \tfrac{1}{2}\Delta\hat{\sigma}^2 + \Delta\hat{\gamma} - \int_{\mathbb{R}} \operatorname{Re}\big(\Delta\psi_n(u)\big) w_\lambda^U(u)\,\mathrm{d}u \\
&\approx \frac{2\pi}{\sqrt{n}} \int_{\mathbb{R}} \Big(-\operatorname{Re}\big(\tfrac{1}{2}\mathcal{F}^{-1} f_{\sigma,U}(-x) \\
&\qquad + \mathcal{F}^{-1} f_{\lambda,U}(-x)\big) + \operatorname{Im}\big(\mathcal{F}^{-1} f_{\gamma,U}(-x)\big)\Big) \varrho(x)\,\mathrm{d}W(x).
\end{aligned}
$$

Therefore, the finite sample variances are given by

$$
\begin{aligned}
s_\gamma^2 &= \frac{4\pi^2}{n} \int_{\mathbb{R}} \Big(-\operatorname{Re}\big(\mathcal{F}^{-1} f_{\sigma,U}(-x)\big) + \operatorname{Im}\big(\mathcal{F}^{-1} f_{\gamma,U}(-x)\big)\Big)^2 \varrho^2(x)\,\mathrm{d}x, && (7.9) \\
s_\lambda^2 &= \frac{4\pi^2}{n} \int_{\mathbb{R}} \Big(-\operatorname{Re}\big(\tfrac{1}{2}\mathcal{F}^{-1} f_{\sigma,U}(-x) + \mathcal{F}^{-1} f_{\lambda,U}(-x)\big) \\
&\qquad + \operatorname{Im}\big(\mathcal{F}^{-1} f_{\gamma,U}(-x)\big)\Big)^2 \varrho^2(x)\,\mathrm{d}x. && (7.10)
\end{aligned}
$$

The estimator $\hat{\nu}(x_0), x_0 \in \mathbb{R}$, in (7.2) involves $\psi_n(u+i)$ instead of $\psi_n(u)$. Hence, the confidence intervals for $\nu(x_0)$ are based on the linearization

$$
\begin{aligned}
\Delta\psi_n(u+i) &:= \psi_n(u+i) - \psi(u+i) \\
&\approx -\frac{u(u+i)}{T\varphi_T(u)} (\mathcal{F}\mathcal{O}_n - \mathcal{F}\mathcal{O})(u+i) \\
&\approx -\frac{1}{\sqrt{n}} \frac{u(u+i)}{T\varphi_T(u)} \int_{\mathbb{R}} e^{iux-x} \varrho(x)\,\mathrm{d}W(x) =: \frac{1}{\sqrt{n}} \mathcal{L}_\nu(u).
\end{aligned}
$$

Defining $f_{\nu,U}(u) := -w_\nu^U(u)u(u+i)/(T\varphi_T(u))$ and writing for brevity $g_U^{(m)}(x_0) := \mathcal{F}^{-1}[u^m w_\nu^U(u)](x_0)$ with $m \in \{0,1,2\}$, the dominating stochastic error term of $\hat{\nu}(x_0)$ is then given by (cf. (4.3))

$$
\begin{aligned}
\Delta\hat{\nu}(x_0) :=& \hat{\nu}(x_0) - \nu(x_0) \\
\approx& \frac{1}{\sqrt{n}} \Big(\frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux_0} \mathcal{L}_\nu(u) w_\nu^U(u)\,\mathrm{d}u \\
&\qquad + \frac{\Delta\hat{\sigma}^2}{2} g_U^{(2)}(x_0) - i\Delta\hat{\gamma} g_U^{(1)}(x_0) + \Delta\hat{\lambda} g_U^{(0)}(x_0)\Big) \\
\approx& \frac{2\pi}{\sqrt{n}} \int_{\mathbb{R}} \Big(\frac{e^{-x}}{2\pi} \mathcal{F}^{-1} f_{\nu,U}(x_0 - x) \\
&\qquad + \operatorname{Re}\big(\mathcal{F}^{-1} f_{\sigma,U}(-x)\big)\big(\tfrac{1}{2} g_U^{(2)}(x_0) + i g_U^{(1)}(x_0) - \tfrac{1}{2} g_U^{(0)}(x_0)\big) \\
&\qquad + \operatorname{Im}\big(\mathcal{F}^{-1} f_{\gamma,U}(-x)\big)\big(-i g_U^{(1)}(x_0) + g_U^{(0)}(x_0)\big) \\
&\qquad - \operatorname{Re}\big(\mathcal{F}^{-1} f_{\lambda,U}(-x)\big) g_U^{(0)}(x_0)\Big) \varrho(x)\,\mathrm{d}W(x),
\end{aligned}
$$

where we note that $g_U^{(0)}, g_U^{(2)}$ are purely real and $g_U^{(1)}$ has only an imaginary part by the symmetry of w_ν^U. Hence, the variance of the linearized stochastic error of $\hat{\nu}(x_0)$ is given by

$$
\begin{aligned}
s^2_{\nu(x_0)} = & \frac{4\pi^2}{n} \int_{\mathbb{R}} \Big(\frac{e^{-x}}{2\pi} \mathcal{F}^{-1} f_{\nu,U}(x_0 - x) \\
& \qquad + \mathrm{Re}\,\big(\mathcal{F}^{-1}\, f_{\sigma,U}(-x)\big)\big(\tfrac{1}{2} g_U^{(2)}(x_0) + i g_U^{(1)}(x_0) - \tfrac{1}{2} g_U^{(0)}(x_0)\big) \\
& \qquad + \mathrm{Im}\,\big(\mathcal{F}^{-1}\, f_{\gamma,U}(-x)\big)\big(- i g_U^{(1)}(x_0) + g_U^{(0)}(x_0)\big) \\
& \qquad - \mathrm{Re}\,\big(\mathcal{F}^{-1}\, f_{\lambda,U}(-x)\big) g_U^{(0)}(x_0)\Big)^2 \varrho^2(x)\,\mathrm{d}x. \qquad (7.11)
\end{aligned}
$$

8 A uniform central limit theorem for deconvolution estimators

In this chapter, the second problem of this thesis is treated, namely the deconvolution problem. Our observations are given by $n \in \mathbb{N}$ independent and identically distributed random variables

$$Y_j = X_j + \varepsilon_j, \qquad j = 1, \dots, n, \tag{8.1}$$

where X_j and ε_j are independent of each other, the distribution of the errors ε_j is supposed to be known and the aim is statistical inference on the distribution of X_j. Let us denote the densities of X_j and ε_j by f_X and f_ε, respectively. The structural similarity of the deconvolution problem and the problem of estimating the characteristic triplet of a Lévy process has been discussed in the introduction and in connection with equation (2.28) in Section 2.2. We consider the case of ordinary smooth errors, which means that the characteristic function φ_ε of the errors ε_j decays with polynomial rate, determining the ill–posedness of the inverse problem. We prove a uniform central limit theorem for kernel estimators of the distribution function of X_j in the setting of $\sqrt{n}$ convergence rates. More precisely, the theorem does not only include the estimation of the distribution function, but covers estimating translation classes of linear functionals of the density f_X. These results will appear in the paper Söhl and Trabs (2012a), where in addition lower bounds for the estimators are studied. In the paper, we show that the used estimators are optimal in the sense of semiparametric efficiency.

The classical Donsker theorem plays a central role in statistics. Let $X_1, \dots, X_n$ be an independent, identically distributed sample from a distribution function F on the real line and define the empirical distribution function by $F_n(t) := \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{X_j \leqslant t\}}$. The classical Donsker theorem states that $\sqrt{n}(F_n - F)$ converges in law to $\mathbb{G}$ in the space of bounded functions $\ell^\infty(\mathbb{R})$, where $\mathbb{G}$ is obtained by the composition of the distribution function F and a standard Brownian bridge. In the deconvolution model (8.1) the random variables $X_1, \dots, X_n$ are not observed, but only $Y_1, \dots, Y_n$, which contain an additive error. In this model we consider estimators $\widehat{\vartheta}_t$ of linear functionals $t \mapsto \vartheta_t := \int \zeta(x-t) f_X(x) \, \mathrm{d}x$, where the special case $\zeta := \mathbf{1}_{(-\infty,0]}$ leads to the estimation of the distribution function F. Our Donsker theorem states that $\sqrt{n}(\widehat{\vartheta}_t - \vartheta_t)_{t \in \mathbb{R}}$ converges in law to a centered Gaussian Borel random variable $\mathbb{G}$ in $\ell^\infty(\mathbb{R})$. This generalization allows to consider functionals ϑ_t as long as the smoothness of ζ in an L^2–Sobolev sense compensates the ill–posedness of the problem.

In order to show the uniform central limit theorem in the deconvolution problem, we prove that the empirical process $\sqrt{n}(\mathbb{P}_n - \mathbb{P})$ is tight in the space of bounded mappings acting on the class

$$\mathcal{G} := \{\mathcal{F}^{-1}[1/\varphi_\varepsilon(-\bullet)] * \zeta_t \,|\, t \in \mathbb{R}\}, \qquad \zeta_t := \zeta(\bullet - t),$$

where $\mathbb{P}$ and $\mathbb{P}_n = \frac{1}{n}\sum_{j=1}^n \delta_{Y_j}$ denote the true and the empirical probability measure of the observations Y_j, respectively. Since $\mathcal{G}$ may consist of translates of an unbounded function, this is in general not a Donsker class. Nevertheless, Radulović and Wegkamp (2000) have observed that a smoothed empirical processes might converge even when the unsmoothed process does not. Giné and Nickl (2008) have further developed these ideas and have shown uniform central limit theorems for kernel density estimators. Nickl and Reiß (2012) used smoothed empirical processes in the inverse problem of estimating the generalized distribution function of Lévy measures.

Additionally to techniques of smoothed empirical processes, our proofs rely on the Fourier multiplier property of the underlying deconvolution operator $\mathcal{F}^{-1}[1/\varphi_\varepsilon]$, which is related to pseudo–differential operators as noted in the Lévy process setting by Nickl and Reiß (2012) and in the deconvolution context by Schmidt–Hieber et al. (2012). Important for our proofs are the mapping properties of $\mathcal{F}^{-1}[1/\varphi_\varepsilon]$ on Besov spaces.

This chapter is organized as follows: In Section 8.1, we define the estimator. In Section 8.2, we formulate the uniform central limit theorem, which we discuss in Section 8.3. Useful properties of the deconvolution operator are collected in Section 8.4. The proof of the uniform central limit theorem is divided into proving the convergence of the finite dimensional distributions in Section 8.5 and tightness in Section 8.6. Definitions and properties of the function spaces used in this chapter are contained in Section 8.7.

8.1 The estimator

According to the observation scheme (8.1), Y_j are distributed with density $f_Y = f_X * f_\varepsilon$ determining the probability measure $\mathbb{P}$. Let φ and φ_ε be the characteristic functions of the Y_j and of the ε_j, respectively. We define the Fourier transform by $\mathcal{F} f(u) := \int e^{iux} f(x)\,\mathrm{d}x, u \in \mathbb{R}$. On the assumption that φ_ε does not vanish, we obtain $\mathcal{F} f_X(u) = \varphi(u)/\varphi_\varepsilon(u)$. For two different functions f_X and $\overline{f_X}$ we denote by φ and $\overline{\varphi}$, respectively, the characteristic functions of the corresponding observations. We have

$$\mathcal{F}[\overline{f_X} - f_X](u) = \frac{\overline{\varphi}(u) - \varphi(u)}{\varphi_\varepsilon(u)}. \tag{8.2}$$

For ζ to be specified later and recalling $\zeta_t = \zeta(\bullet - t)$, our aim is to estimate functionals of the form

$$\vartheta_t := \langle \zeta_t, f_X \rangle = \int \zeta_t(x) f_X(x)\,\mathrm{d}x. \tag{8.3}$$

The characteristic function φ of $\mathbb{P}$ can be estimated by its empirical version $\varphi_n(u) = \frac{1}{n}\sum_{j=1}^n e^{iuY_j}, u \in \mathbb{R}$. Equation (8.2) shows that the difference between two characteristic functions $\overline{\varphi}$ and φ will be divided by the decaying function φ_ε. So we apply a spectral cut–off by multiplying with the Fourier transform of a band–limited kernel. The natural estimator of the functional ϑ_t is given by

$$\widehat{\vartheta}_t := \int \zeta_t(x)\,\mathcal{F}^{-1}\Big[\mathcal{F}K_h\frac{\varphi_n}{\varphi_\varepsilon}\Big](x)\,\mathrm{d}x, \tag{8.4}$$

where K is a band–limited kernel, $h > 0$ the bandwidth and we have written as usual $K_h(x) = h^{-1}K(x/h)$. Choosing $\mathcal{F}K = \mathbf{1}_{[-\kappa,\kappa]}$ for some $\kappa > 0$ leads to the estimator proposed by Butucea and Comte (2009). Throughout, we suppose that

(i) $K \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ is symmetric and band–limited with $\operatorname{supp}(\mathcal{F}K) \subseteq [-1, 1]$,

(ii) for $l = 1, \dots, L$

$$\int K = 1, \quad \int x^l K(x)\,\mathrm{d}x = 0, \quad \int |x^{L+1}K(x)|\,\mathrm{d}x < \infty \quad \text{and} \tag{8.5}$$

(iii) $K \in C^1(\mathbb{R})$ satisfies, denoting $\langle x\rangle := (1 + x^2)^{1/2}$,

$$|K(x)| + |K'(x)| \lesssim \langle x\rangle^{-2}. \tag{8.6}$$

Throughout, we write $A_p \lesssim B_p$ if there exists a constant $C > 0$ independent of the parameter p such that $A_p \leqslant CB_p$. If $A_p \lesssim B_p$ and $B_p \lesssim A_p$, we write $A_p \sim B_p$. Examples of such kernels can be obtained by taking $\mathcal{F}K$ to be a symmetric function in $C^\infty(\mathbb{R})$ which is supported in $[-1, 1]$ and constant to one in a neighborhood of zero. The resulting kernels are called flat top kernels and were used in deconvolution problems, for example, by Bissantz et al. (2007).

8.2 Statement of the theorem

Given a function ζ specified later, our aim is to show a uniform central limit theorem for the estimator over the class of translations ζ_t, $t \in \mathbb{R}$. In view of the classical Donsker theorem in a model without additive errors, where no assumptions on the smoothness of the distribution are needed, we want to assume as less smoothness of f_X as possible still guaranteeing $\sqrt{n}$–rates. For some $\delta > 0$ the following assumptions on the density f_X will be needed:

Assumption 8.1.

(i) Let f_X be bounded and assume the moment condition $\int |x|^{2+\delta} f_X(x)\,\mathrm{d}x < \infty$.

(ii) Assume $f_X \in H^\alpha(\mathbb{R})$ that is the density has Sobolev smoothness of order $\alpha \geqslant 0$.

We refer to the appendix for an exact definition of the Sobolev space $H^\alpha(\mathbb{R})$. Boundedness of the observation density f_Y follows immediately from point (i) of the assumption since $\|f_Y\|_\infty \leqslant \|f_X\|_\infty \|f_\varepsilon\|_{L^1} < \infty$. In addition to the smoothness of f_X, the smoothness of ζ will be crucial. We assume for $\gamma_s, \gamma_c > 0$

$$\zeta \in Z^{\gamma_s,\gamma_c} := \Big\{\zeta = \zeta^c + \zeta^s \Big| \zeta^s \in H^{\gamma_s}(\mathbb{R}) \text{ is compactly supported as well as } \langle x\rangle^\tau(\zeta^c(x) - a(x)) \in H^{\gamma_c}(\mathbb{R}) \text{ for some } \tau > 0 \text{ and some } a \in C^\infty(\mathbb{R}) \text{ such that } a' \text{ is compactly supported}\Big\} \tag{8.7}$$

and write for $\zeta \in Z^{\gamma_s,\gamma_c}$ with a given decomposition $\zeta = \zeta^s + \zeta^c$

$$\|\zeta\|_{Z^{\gamma_s,\gamma_c}} := \|\zeta^s\|_{H^{\gamma_s}} + \big\|\tfrac{1}{ix+1}\zeta^c(x)\big\|_{H^{\gamma_c}},$$

which is finite since the second term $\|\frac{1}{ix+1}\zeta^c(x)\|_{H^{\gamma_c}}$ can be bounded by $\|\frac{a(x)}{ix+1}\|_{H^{\gamma_c}} + \|\frac{1}{(ix+1)\langle x\rangle^\tau}\|_{C^s}\|\langle x\rangle^\tau(\zeta^c(x) - a(x))\|_{H^{\gamma_c}} < \infty$ for any $s > \gamma_c$. Let us give two examples for ζ and corresponding γ_s, γ_c.

Example 8.2. To estimate the distribution function of X_j, one has to consider translations of the indicator function $\mathbf{1}_{(-\infty,0]}(x)$, $x \in \mathbb{R}$. Let a be a monotone decreasing $C^\infty(\mathbb{R})$ function, which is for some $M > 0$ equal to zero for all $x \geqslant M$ and equal to one for all $x \leqslant -M$. We define $\zeta^s := \mathbf{1}_{(-\infty,0]} - a$ and $\zeta^c := a$. From the bounded variation of ζ^s follows $\zeta^s \in B^1_{1,\infty}(\mathbb{R}) \subseteq H^{\gamma_s}(\mathbb{R})$ for any $\gamma_s < 1/2$ by Besov smoothness of bounded variation functions (8.38) as well as by the Besov space embeddings (8.33) and (8.34). Since $a \in C^\infty(\mathbb{R})$ and a' is compactly supported, the condition on ζ^c is satisfied for any $\gamma_c > 0$. Hence, $\mathbf{1}_{(-\infty,t]} \in Z^{\gamma_s,\gamma_c}$ if $\gamma_s < 1/2$. On the other hand, this cannot hold for $\gamma_s > 1/2$ since $H^{\gamma_s}(\mathbb{R}) \subseteq C^0(\mathbb{R})$ by Sobolev's embedding theorem or by (8.32), (8.33) and (8.34).

Example 8.3. In the context of M–estimation (or Z–estimation) the root of the equation

$$\langle \zeta(\cdot - t), f_X\rangle = 0$$

is used for inference, e.g., on the location of the distribution of X_j. A popular example in robust statistics is the Huber estimator where $\zeta(x) = h_K(x) := ((-K) \vee x) \wedge K$ for some $K > 0$. In that case a similar decomposition as in Example 8.2 shows $h_K \in Z^{\gamma_s,\gamma_c}$ for any $\gamma_s < 3/2$.

The ill–posedness of the problem is determined by the decay of the characteristic function of the errors. More precisely, we suppose

Assumption 8.4. Let the error distribution satisfy

(i) $\int |x|^{2+\delta} f_\varepsilon(x)\,\mathrm{d}x < \infty$ thus φ_ε is twice continuously differentiable and

(ii) $|(\varphi_\varepsilon^{-1})'(u)| \lesssim \langle u\rangle^{\beta-1}$ for some $\beta > 0$, in particular $|\varphi_\varepsilon^{-1}(u)| \lesssim \langle u\rangle^\beta, u \in \mathbb{R}$.

Throughout, we write $\varphi_\varepsilon^{-1} = 1/\varphi_\varepsilon$. The Assumption (ii) on the distribution of the errors is similar to the classical decay assumption by Fan (1991a) and it is fulfilled for many ordinary smooth error laws such as gamma or Laplace distributions as discussed below. Assumption 8.4(ii) implies that φ_ε^{-1} is a Fourier multiplier on Besov spaces so that

$$B^s_{p,q}(\mathbb{R}) \ni f \mapsto \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}f] \in B^{s-\beta}_{p,q}(\mathbb{R})$$

for $p,q \in [1,\infty], s \in \mathbb{R}$, is a continuous linear map, which is essential in our proofs, compare Lemma 8.8. In the same spirit Schmidt–Hieber et al. (2012) discuss the behavior of the deconvolution operator as pseudo–differential operator. We define

$$g_t := \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta_t \quad \text{and} \quad \mathcal{G} = \{g_t | t \in \mathbb{R}\}. \tag{8.8}$$

Note that in general g_t may only exist in a distributional sense, but on Assumption 8.4 and for $\zeta \in Z^{\gamma_s,\gamma_c}$ it can be rigorously interpreted by (see (8.17))

$$\begin{aligned} g_0(x) = &\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}\zeta^s(u)](x) \\ &+ (1+ix)\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)](x) \\ &+ \mathcal{F}^{-1}[(\varphi_\varepsilon^{-1})'(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)](x), \end{aligned}$$

which indicates why we have imposed an assumption on $(\varphi_\varepsilon^{-1})'$ and have defined $\|\bullet\|_{Z^{\gamma_s,\gamma_c}}$ as above.

It will turn out that $\mathcal{G}$ is $\mathbb{P}$–pregaussian, but not Donsker in general. Denoting by $\lfloor\alpha\rfloor$ the largest integer smaller or equal to α and defining convergence in law on $\ell^\infty(\mathbb{R})$ as Dudley (1999, p. 94), we state our main result

Theorem 8.5. *Grant Assumptions 8.1 and 8.4 as well as $\zeta \in Z^{\gamma_s,\gamma_c}$ with $\gamma_s > \beta, \gamma_c > (1/2 \vee \alpha) + \gamma_s$ and $\alpha + 3\gamma_s > 2\beta + 1$. Furthermore, let the kernel K satisfy* (8.5) *with $L = \lfloor \alpha + \gamma_s \rfloor$. Let $h_n^{2\alpha+2\gamma_s} n \to 0$ and if $\gamma_s \leqslant \beta + 1/2$ let in addition $h_n^\rho n \to \infty$ for some $\rho > 4\beta - 4\gamma_s + 2$, then*

$$\sqrt{n}(\widehat{\vartheta}_t - \vartheta_t)_{t\in\mathbb{R}} \xrightarrow{\mathcal{L}} \mathbb{G} \quad \text{in } \ell^\infty(\mathbb{R})$$

as $n \to \infty$, where $\mathbb{G}$ is a centered Gaussian Borel random variable in $\ell^\infty(\mathbb{R})$ with covariance function given by

$$\Sigma_{s,t} := \int g_s(x) g_t(x)\,\mathbb{P}(\,\mathrm{d}x) - \vartheta_s\vartheta_t$$

for g_s, g_t defined in (8.8) *and $s,t \in \mathbb{R}$.*

We illustrate the range of this theorem by the following examples.

Example 8.6. For estimating the distribution function Assumption 8.4 needs to be fulfilled for some $\beta < 1/2$ owing to the condition $\gamma_s > \beta$. This is fulfilled by the gamma distribution $\Gamma(\beta,\eta)$ with $\beta \in (0,1/2)$ and $\eta \in (0,\infty)$, that is

$$f_\varepsilon(x) := \gamma_{\beta,\eta}(x) := \frac{1}{\Gamma(\beta)\eta^\beta} x^{\beta-1} e^{-x/\eta} \mathbf{1}_{[0,\infty)}(x), \quad x \in \mathbb{R},$$

and $\varphi_\varepsilon(u) = (1 - i\eta u)^{-\beta}$, $u \in \mathbb{R}$.

For the Huber estimator from Example 8.3 we required $\beta < 3/2$, which holds, for instance, for the chi–squared distribution with one or two degrees of freedom or for the exponential distribution.

Example 8.7. Butucea and Comte (2009) studied the case $\beta > 1$ and derived $\sqrt{n}$–rates for $\gamma_s > \beta$ in our notation. In particular, they considered supersmooth ζ, that is $\mathcal{F}\zeta$ decays exponentially. In this case $\zeta \in H^s(\mathbb{R})$ for any $s \in \mathbb{N}$. Requiring the slightly stronger assumption that $\langle x\rangle^\tau \zeta(x) \in H^s(\mathbb{R})$ for some arbitrary small $\tau > 0$ and for all $s \in \mathbb{N}$ we can choose $\zeta^c := \zeta$ and $\zeta^s := 0$. Then β can be taken arbitrary large such that all gamma distributions, the Laplace distributions and convolutions of them can be chosen as error distributions.

8.3 Discussion

To have $\sqrt{n}$–rates we suppose $\gamma_s > \beta$ in Theorem 8.5, which means that the smoothness of the functionals compensates the ill–posedness of the problem. This condition is natural in view of the abstract analysis in terms of Hilbert scales by Goldenshluger and Pereverzev (2003), who obtain the minimax rate $n^{-(\alpha+\gamma_s)/(2\alpha+2\beta)} \vee n^{-1/2}$ in our notation. As a consequence of the condition on γ_s and γ_c we can bound the stochastic error term of the estimator $\widehat{\vartheta}_t$ uniformly in $h \in (0,1)$. The bias term is of order $h^{\alpha+\gamma_s}$.

For $\gamma_s > \beta + 1/2$ the class $\mathcal{G}$ is a Donsker class. In this case the only condition on the bandwidth is that the bias tends faster than $n^{-1/2}$ to zero. In the interesting but involved case $\gamma_s \in (\beta, \beta + 1/2]$, the class $\mathcal{G}$ will in general not be a Donsker class. Estimating the distribution function as in Example 8.2 belongs to this case. In order to see that $\mathcal{G}$ is in general not a Donsker class, let the error distribution be given by $f_\varepsilon = \gamma_{\beta,\eta}(-\bullet)$ and $\zeta = \gamma_{\sigma,\eta}$ with $\sigma \in (\gamma_s + 1/2, \beta + 1)$. Then g_t equals $\gamma_{\sigma-\beta,\eta} * \delta_t$. For the shape parameter holds $\sigma - \beta \in (1/2, 1)$ and thus g_t is an $L^2(\mathbb{R})$–function unbounded at t. The Lebesgue density of $\mathbb{P}$ is bounded by Assumption 8.1(i). Hence, $\mathcal{G}$ consists of all translates of an unbounded function and thus cannot be Donsker, cf. Theorem 7 by Nickl (2006).

Therefore, for $\gamma_s \in (\beta, \beta + 1/2]$ smoothed empirical processes are necessary, especially we need to ensure enough smoothing to be able to obtain a uniform central limit theorem. The bandwidth cannot tend too fast to zero, more precisely we require $h_n^\rho n \to \infty$ as $n \to \infty$ for some ρ with $\rho > 4\beta - 4\gamma_s + 2$. In combination with the bias condition $h_n^{2\alpha+2\gamma_s} n \to 0$ as $n \to \infty$ we obtain necessarily $\alpha + \gamma_s > 2\beta - 2\gamma_s + 1$ leading to the assumption in the theorem. Since $2\alpha + 2\gamma_s > \alpha + 2\beta - \gamma_s + 1 > 4\beta - 4\gamma_s + 2$ we can always choose $h_n \sim n^{-1/(\alpha+2\beta-\gamma_s+1)}$. In contrast to Butucea and Comte (2009); Dattner et al. (2011); Fan (1991b) our choice of the bandwidth h_n is not determined by the bias–variance trade–off, but rather by the amount of smoothing necessary to obtain a uniform central limit theorem. The classical bandwidth $h_n \sim n^{-1/(2\alpha+2\beta)}$ is optimal for estimating the density in the sense that it achieves the minimax rate with respect to the mean integrated squared error (MISE), compare Fan (1991b) who assumes Hölder smoothness of f_X instead of L^2–Sobolev smoothness. For this choice the bias condition $h_n^{2\alpha+2\gamma_s} n \to 0$ is satisfied.

If $\gamma_s \leqslant \beta + 1/2$ the classical bandwidth satisfies the additional minimal smoothness condition in the case of estimating the distribution function with mild conditions on f_X. It suffices for example that f_X is of bounded variation. Then α and γ_s can be chosen large enough in $(0, 1/2)$ such that $2\alpha + 2\beta > 4\beta - 4\gamma_s + 2$ and the classical bandwidth satisfies the conditions of the theorem. Whenever the classical bandwidth $h_n \sim n^{-1/(2\alpha+2\beta)}$ satisfies the conditions of Theorem 8.5, then the corresponding density estimator is a 'plug–in' estimator in the sense of Bickel and Ritov (2003) meaning that the density is estimated rate optimal for the MISE and the estimators of the functionals converge uniformly over $t \in \mathbb{R}$.

The smoothness condition on the density f_X is then a consequence of the given choice of h_n together with the classical bias estimate for kernel estimators. As we have seen in Example 8.2 for estimating the distribution function, we have $\zeta = \mathbf{1}_{(-\infty,0]} \in Z^{\gamma_s,\gamma_c}$ with $\gamma_s < 1/2$ arbitrary close to 1/2. In the classical Donsker theorem which corresponds to the case $\beta \to 0$ the condition $\alpha + 3\gamma_s > 2\beta + 1$ would simplify to $\alpha > -1/2$. However, we suppose f_X to be bounded, which leads to much clearer proofs, and thus $f_X \in H^0(\mathbb{R})$ is automatically satisfied. Assumption 8.1 allows to focus on the interplay between the functional ζ and the deconvolution operator $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}]$. Nickl and Reiß (2012) have studied the case of unbounded densities, which is necessary in the Lévy process setup, but considered $\zeta_t = \mathbf{1}_{(-\infty,t]}$ only. The class Z^{γ_s,γ_c} is defined by L^2–Sobolev conditions so that bounded variation arguments for ζ have to be avoided in the proofs.

An interesting aspect is the following: If we restrict the uniform convergence to $(\zeta_t)_{t\in T}$ for some compact set $T \subseteq \mathbb{R}$, it is sufficient to assume $\frac{1}{ix+1}\zeta^c \in H^{\gamma_c}(\mathbb{R})$ instead of requiring $(1 \vee |x|^\tau)(\zeta^c(x) - a(x)) \in H^{\gamma_c}(\mathbb{R})$ for some $\tau > 0$ and a function $a \in C^\infty(\mathbb{R})$ such that a' is compactly supported as done in Z^{γ_s,γ_c}. In particular, slowly growing ζ would be allowed. The stronger condition in the definition of Z^{γ_s,γ_c} is only needed to ensure polynomial covering numbers of $\{g_t | t \in T\}$ for $T \subseteq \mathbb{R}$ unbounded (cf. Theorem 8.10 below).

As a corollary of Theorem 8.5 we can weaken Assumption 8.4(ii). If the characteristic function of the errors ε is given by $\widetilde{\varphi}_\varepsilon = \varphi_\varepsilon \psi$ where φ_ε satisfies Assumption 8.4(ii) and there is a Schwartz distribution $\nu \in \mathscr{S}'(\mathbb{R})$ such that $\mathcal{F}\nu = \psi^{-1}$ and $\nu * \zeta \in Z^{\gamma_s,\gamma_c}$ for $\zeta \in Z^{\gamma_s,\gamma_c}$, then for $t \in \mathbb{R}$

$$\mathcal{F}^{-1}[\widetilde{\varphi}_\varepsilon^{-1}] * \zeta(\bullet - t) = \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * (\nu * \zeta)(\bullet - t)$$

and thus we can proceed as before. For instance, for translated errors $f_\varepsilon * \delta_\mu$ with $\mu \neq 0$, the distribution ν would be given by $\delta_{-\mu}$. Thus even if $\widetilde{\varphi}_\varepsilon$ does not satisfy Assumption 8.4(ii) the statement of Theorem 8.5 holds true if we can write $\widetilde{\varphi}_\varepsilon(u) = \varphi_\varepsilon(u)e^{i\mu u}$ with φ_ε satisfying Assumption 8.4(ii). The assumption by Fan and Liu (1997) for the asymptotic normality in the case of ordinary smooth deconvolution are weaker than the assumptions by Fan (1991a) since they allow exactly for this additional factor $e^{i\mu u}$ in the assumptions on the characteristic function of the error.

As for the classical Donsker theorem the Donsker theorem for deconvolution estimators has many different applications, the most obvious being the construction of confidence sets. Further Donsker theorems may be obtained by applying the functional delta method to Hadamard differentiable maps.

In order to construct confidence intervals or joint confidence sets at finitely many points we take m different points $t_1, \ldots, t_m \in \mathbb{R}$. Let the assumptions of Theorem 8.5 be fulfilled. Then $g_t \in L^2(\mathbb{P})$ for all $t \in \mathbb{R}$ and $\hat{\Sigma}_{s,t} = \int g_s(x) g_t(x) \, \mathbb{P}_n(\,\mathrm{d}x) - \widehat{\vartheta}_s \widehat{\vartheta}_t$ is a consistent estimator of the covariance $\Sigma_{s,t} = \int g_s(x) g_t(x) \, \mathbb{P}(\,\mathrm{d}x) - \vartheta_s \vartheta_t$ by the weak law of large numbers. The spectral theorem yields an orthogonal matrix O and a diagonal matrix D such that the covariance matrix Σ corresponding to the points $t_1, \ldots, t_m \in \mathbb{R}$ can be written as $\Sigma = ODO^\top$. We assume that the determinant of Σ is nonzero and thus the diagonal entries of D are positive. By the convergence of the finite dimensional distributions we have

$$\sqrt{n} \begin{pmatrix} \widehat{\vartheta}_{t_1} - \vartheta_{t_1} \\ \vdots \\ \widehat{\vartheta}_{t_m} - \vartheta_{t_m} \end{pmatrix} \xrightarrow{d} X, \tag{8.9}$$

where $X \sim N(0, \Sigma)$. The distribution of X is the same as the distribution of $OD^{1/2}Y$ for a random vector Y of m independent standard normal random variables. By the continuous mapping theorem we can apply the map $(OD^{1/2})^{-1}$ to (8.9). Let k_η denote the $(1-\eta)$–quantile of the chi–squared distribution χ^2_m of m degrees of freedom. We define

$$A_n := (\widehat{\vartheta}_{t_1}, \ldots, \widehat{\vartheta}_{t_m})^\top + n^{-1/2} OD^{1/2} \left\{ x \in \mathbb{R}^m \mid \|x\| \leqslant k_\eta \right\},$$

where $\|x\|$ denotes the Euclidean norm. Then we have the following confidence statement

$$\lim_{n \to \infty} \mathbb{P}((\vartheta_{t_1}, \ldots, \vartheta_{t_m}) \in A_n) = 1 - \eta.$$

The special case $m = 1$ yields confidence intervals.

Let us illustrate the construction of confidence bands. By the continuous mapping theorem we infer

$$\sup_{t \in \mathbb{R}} \sqrt{n} |\widehat{\vartheta}_t - \vartheta_t| \xrightarrow{\mathcal{L}} \sup_{t \in \mathbb{R}} |\mathbb{G}(t)|.$$

The construction of confidence bands reduces now to knowledge about the distribution of the supremum of $\mathbb{G}$. Suprema of Gaussian processes are well studied and information about their distribution can be either obtained from theoretical considerations as in van der Vaart and Wellner (1996, App. A.2) or from Monte Carlo simulations. Let $q_{1-\eta}$ be the $(1-\eta)$–quantile of $\sup_{t \in \mathbb{R}} |\mathbb{G}(t)|$ that is $\mathbb{P}(\sup_{t \in \mathbb{R}} |\mathbb{G}(t)| \leqslant q_{1-\eta}) = 1 - \eta$. Then

$$\lim_{n \to \infty} \mathbb{P} \left(\vartheta_t \in [\widehat{\vartheta}_t - q_{1-\eta} n^{-1/2}, \widehat{\vartheta}_t + q_{1-\eta} n^{-1/2}] \text{ for all } t \in \mathbb{R} \right) = 1 - \eta$$

and thus the intervals $[\widehat{\vartheta}_t - q_{1-\eta} n^{-1/2}, \widehat{\vartheta}_t + q_{1-\eta} n^{-1/2}]$ define a confidence band.

As often in the literature we assume the error distribution to be known. In a more realistic setting where an independent sample of pure noise observations is observable, Neumann (1997) constructs a density estimator for unknown errors. In this case the deconvolution operator $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}]$ needs to be estimated leading to an inverse problem with an error in the operator. A Donsker theorem in this situation is related to the Lévy process setting of Nickl and Reiß (2012) since their estimator is based on the deconvolution with the unknown marginal distribution of the process. Another interesting aspect is the similarity of the deconvolution problem and the errors–in–variables regression, see for example Fan and Truong (1993) and the references therein. The corresponding statement to our result would be a uniform central limit theorem for linear functionals of the regression function, where similar methods may be applied.

8.4 The deconvolution operator

In this section, we provide an auxiliary lemma, which describes the properties of the deconvolution operator $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}]$.

Lemma 8.8. *Grant Assumption 8.4.*

(i) For all $s \in \mathbb{R}, p, q \in [1,\infty]$ the deconvolution operator $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)]$ is a Fourier multiplier from $B^s_{p,q}(\mathbb{R})$ to $B^{s-\beta}_{p,q}(\mathbb{R})$, that is the linear map

$$B^s_{p,q}(\mathbb{R}) \to B^{s-\beta}_{p,q}(\mathbb{R}), f \mapsto \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F} f]$$

is bounded.

(ii) For any integer m strictly larger then β we have $\mathcal{F}^{-1}[(1+iu)^{-m}\varphi_\varepsilon^{-1}] \in L^1(\mathbb{R})$ and if $m > \beta + 1/2$ we also have $\mathcal{F}^{-1}[(1+iu)^{-m}\varphi_\varepsilon^{-1}] \in L^2(\mathbb{R})$.

(iii) Let $\beta^+ > \beta$ and $f, g \in H^{\beta^+}(\mathbb{R})$. Then

$$\int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * f) g = \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * g) f. \tag{8.10}$$

Using the kernel K, this equality extends to functions $g \in L^2(\mathbb{R}) \cup L^\infty(\mathbb{R})$ and finite Borel measures μ:

$$\int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}\,\mathcal{F} K_h] * \mu) g = \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F} K_h] * g)\,\mathrm{d}\mu. \tag{8.11}$$

Proof.

(i) Analogously to the proof by Nickl and Reiß (2012), we deduce from Corollary 4.11 of Girardi and Weis (2003) that $(1+iu)^{-\beta}\varphi_\varepsilon^{-1}(-u)$ is a Fourier multiplier on $B^s_{p,q}$ by Assumption 8.4(ii). It remains to note that $j : B^s_{p,q}(\mathbb{R}) \to B^{s-\beta}_{p,q}(\mathbb{R}), f \mapsto \mathcal{F}^{-1}[(1+iu)^\beta\,\mathcal{F} f]$ is a linear isomorphism (Triebel, 2010, Thm. 2.3.8).

(ii) Since the gamma density $\gamma_{1,1}$ is of bounded variation, it is contained in $B^1_{1,\infty}(\mathbb{R})$ by (8.38). Using the isomorphism j from (i), we deduce $\gamma_{m,1} \in B^m_{1,\infty}(\mathbb{R})$ and thus by Besov embeddings (8.34) and (8.31)

$$\mathcal{F}^{-1}[(1+iu)^{-m}\varphi_\varepsilon^{-1}] \in B^{m-\beta}_{1,\infty}(\mathbb{R}) \subseteq B^0_{1,1}(\mathbb{R}) \subseteq L^1(\mathbb{R}).$$

If $m-\beta > 1/2$, we can apply the embedding $B^{m-\beta}_{1,\infty}(\mathbb{R}) \subseteq B^{m-\beta-1/2}_{2,\infty}(\mathbb{R}) \subseteq L^2(\mathbb{R})$.

(iii) For $f \in H^{\beta+}(\mathbb{R})$, (i) and the Besov embeddings (8.31), (8.33) and (8.34) yield

$$\|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * f\|_{L^2} \lesssim \|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * f\|_{B^0_{2,1}} \lesssim \|f\|_{B^\beta_{2,1}} \lesssim \|f\|_{H^{\beta+}} < \infty.$$

Therefore, it follows by Plancherel's equality

$$\begin{aligned}\int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * f)(x) g(x)\,\mathrm{d}x &= \frac{1}{2\pi}\int \varphi_\varepsilon^{-1}(-u)\,\mathcal{F} f(-u)\,\mathcal{F} g(u)\,\mathrm{d}u \\ &= \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * g)(x) f(x)\,\mathrm{d}x.\end{aligned}$$

To prove the second part of the claim for $g \in L^2(\mathbb{R})$, we note that by Young's inequality

$$\|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}\,\mathcal{F} K_h]\|_{L^2} \leqslant \|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}\mathbf{1}_{[-1/h,1/h]}]\|_{L^2}\|K_h\|_{L^1} < \infty$$

due to the support of $\mathcal{F} K$ and Assumption (8.6) on the decay of K. Since μ is a finite measure and g is bounded, Fubini's theorem yields then

$$\begin{aligned}&\int g(x)(\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}\,\mathcal{F} K_h] * \mu)(x)\,\mathrm{d}x \\ &\quad= \int\int g(x)\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}\,\mathcal{F} K_h](x-y)\mu(\mathrm{d}y)\,\mathrm{d}x \\ &\quad= \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F} K_h] * g)(y)\mu(\mathrm{d}y),\end{aligned}$$

where we have used the symmetry of the kernel. In order to apply Fubini's theorem for the case $g \in L^\infty(\mathbb{R})$, too, we have to show that $\|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}\,\mathcal{F} K_h]\|_{L^1}$ is finite. We replace the indicator function by a function $\chi \in C^\infty(\mathbb{R})$ which equals one on $[-1/h, 1/h]$ and is compactly supported. We estimate

$$\|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}\,\mathcal{F} K_h]\|_{L^1} \leqslant \|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}\chi]\|_{L^1}\|K_h\|_{L^1}. \tag{8.12}$$

Using that $\varphi_\varepsilon^{-1}\chi$ is twice continuously differentiable and has got compact support, we obtain

$$\begin{aligned}\|(1+x^2)\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}\chi](x)\|_\infty &\leqslant \|\mathcal{F}^{-1}[(\mathrm{Id}-\mathrm{D}^2)\varphi_\varepsilon^{-1}\chi](x)\|_\infty \\ &\leqslant \|(\mathrm{Id}-\mathrm{D}^2)\varphi_\varepsilon^{-1}\chi\|_{L^1} < \infty,\end{aligned}$$

where we denote the identity and the differential operator by Id and D, respectively.

This shows that (8.12) is finite. □

8.5 Convergence of the finite dimensional distributions

As usual, we decompose the error into a stochastic error term and a bias term:

$$\begin{aligned}\widehat{\vartheta}_t - \vartheta_t &= \widehat{\vartheta}_t - \mathbb{E}[\widehat{\vartheta}_t] + \mathbb{E}[\widehat{\vartheta}_t] - \vartheta_t \\ &= \int \zeta_t(x)\,\mathcal{F}^{-1}\Big[\mathcal{F}\,K_h\frac{\varphi_n-\varphi}{\varphi_\varepsilon}\Big](x)\,\mathrm{d}x + \int \zeta_t(x)(K_h * f_X(x) - f_X(x))\,\mathrm{d}x.\end{aligned}$$

8.5.1 The bias

The bias term can be estimated by the standard kernel estimator argument. Let us consider the singular and the continuous part of ζ separately. Applying Plancherel's identity and Hölder's inequality, we obtain

$$\begin{aligned}&\Big|\int \zeta_t^s(x)(K_h * f_X(x) - f_X(x))\,\mathrm{d}x\Big| \\ &= \frac{1}{2\pi}\Big|\int \mathcal{F}\,\zeta_t^s(u)(\mathcal{F}\,K(hu)-1)\,\mathcal{F}\,f_X(-u)\,\mathrm{d}u\Big| \\ &\leqslant \|\langle u\rangle^{-(\alpha+\gamma_s)}(\mathcal{F}\,K(hu)-1)\|_\infty \int \langle u\rangle^{\alpha+\gamma_s}|\,\mathcal{F}\,\zeta^s(u)\,\mathcal{F}\,f_X(u)|\,\mathrm{d}u \\ &\leqslant h^{\alpha+\gamma_s}\|u^{-(\alpha+\gamma_s)}(\mathcal{F}\,K(u)-1)\|_\infty\|\zeta^s\|_{H^{\gamma_s}}\|f_X\|_{H^\alpha}.\end{aligned}$$

The term $\|u^{-(\alpha+\gamma_s)}(\mathcal{F}\,K(u)-1)\|_\infty$ is finite using the a Taylor expansion of $\mathcal{F}\,K$ around zero with $(\mathcal{F}\,K)^{(l)}=0$ for $l=1,\dots,\lfloor\alpha+\gamma_s\rfloor$ by the order of the kernel (8.5).

For the smooth part of ζ_t Plancherel's identity yields

$$\begin{aligned}&\Big|\int \zeta_t^c(x)(K_h * f_X - f_X)(x)\,\mathrm{d}x\Big| \\ &= \frac{1}{2\pi}\Big|\int \mathcal{F}[\tfrac{1}{ix+1}\zeta_t^c(x)](\mathrm{Id}+\mathrm{D})\{(\mathcal{F}\,K(hu)-1)\,\mathcal{F}\,f_X(-u)\}\,\mathrm{d}u\Big| \\ &\leqslant \int |\,\mathcal{F}[\tfrac{1}{ix+1}\zeta_t^c(x)](\mathcal{F}\,K(hu)-1+h\,\mathcal{F}[ixK](hu))\,\mathcal{F}\,f_X(-u)|\,\mathrm{d}u \\ &\quad - \int |\,\mathcal{F}[\tfrac{1}{ix+1}\zeta_t^c(x)](\mathcal{F}\,K(hu)-1)\,\mathcal{F}[ixf_X](-u)|\,\mathrm{d}u.\end{aligned}$$

The first term can be estimated as before and for the second term we note that $xf_X(x)\in L^2(\mathbb{R})=H^0(\mathbb{R})$ by Assumption 8.1(i) such that the additional smoothness of $\frac{1}{ix+1}\zeta^c(x)$ yields the right order. Therefore, we have $|\,\mathbb{E}[\widehat{\vartheta}_t]-\vartheta_t|\lesssim h^{\alpha+\gamma_s}$ and thus by the choice of h, the bias term is of order $o(n^{-1/2})$.

8.5.2 The stochastic error

We notice that $\|\zeta^c - a\|_{H^{\gamma_c}} \lesssim \|\langle x\rangle^{-\tau}\|_{C^s}\|\langle x\rangle^{\tau}(\zeta^c(x)-a(x))\|_{H^{\gamma_c}} < \infty$ for any $s > \gamma_c$, where we used the pointwise multiplier property (8.35) as well as the Besov embeddings (8.34) and (8.32). We have $\zeta^s \in L^2$ and by (8.31), (8.33) and (8.34)

$$\|\zeta^c\|_\infty \leqslant \|a\|_\infty + \|\zeta^c - a\|_\infty \leqslant \|a\|_\infty + \|\zeta^c - a\|_{H^{\gamma_c}} < \infty,$$

since $\gamma_c > 1/2$. Consequently, we can apply the smoothed adjoint equality (8.11) and obtain for the stochastic error term

$$\begin{aligned}&\int \zeta_t(x)\,\mathcal{F}^{-1}\Big[\mathcal{F} K_h \frac{\varphi_n-\varphi}{\varphi_\varepsilon}\Big](x)\,\mathrm{d}x\\ &\quad= \int \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F} K_h] * \zeta_t(x)(\mathbb{P}_n-\mathbb{P})(\,\mathrm{d}x).\end{aligned}\tag{8.13}$$

Therefore, it suffices for the convergence of the finite dimensional distributions to bound the term

$$\sup_{h\in(0,1)}\int \Big|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F} K_h] * \zeta(x)\Big|^{2+\delta}\,\mathbb{P}(\,\mathrm{d}x),\tag{8.14}$$

for any function $\zeta \in Z^{\gamma_s,\gamma_c}$. Then the stochastic error term converges in distribution to a normal random variable by the central limit theorem under the Lyapunov condition (i.e., Klenke, 2007, Thm. 15.43 together with Lem. 15.41). Finally, the Cramér–Wold device yields the convergence of the finite dimensional distributions in Theorem 8.5.

First, note that the moment conditions in Assumptions 8.1 and 8.4 and the estimate

$$|x|^p f_Y(x) \leqslant \int |x-y+y|^p f_X(x-y) f_\varepsilon(y)\,\mathrm{d}y \lesssim (|y|^p f_X) * f_\varepsilon + f_X * (|y|^p f_\varepsilon),$$

for $x \in \mathbb{R}$, $p \geqslant 1$, yield finite $(2+\delta)$th moments for $\mathbb{P}$ since

$$\int |x|^{2+\delta} f_Y(x)\,\mathrm{d}x \lesssim \||x|^{2+\delta} f_X\|_{L^1}\|f_\varepsilon\|_{L^1} + \|f_X\|_{L^1}\||x|^{2+\delta} f_\varepsilon\|_{L^1} < \infty.\tag{8.15}$$

To estimate (8.14), we rewrite

$$\begin{aligned}\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta^c(x) &= \mathcal{F}^{-1}\left[\varphi_\varepsilon^{-1}(-u)(\mathrm{Id}+\mathrm{D})\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)\right](x)\\ &= \mathcal{F}^{-1}\left[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)\right](x)\\ &\quad+ \mathcal{F}^{-1}\left[\varphi_\varepsilon^{-1}(-u)\big(\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)]\big)'(u)\right](x)\\ &= (1+ix)\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)](x)\\ &\quad+ \mathcal{F}^{-1}[(\varphi_\varepsilon^{-1})'(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)](x),\end{aligned}\tag{8.16}$$

owing to the product rule for differentiation. Hence,

$$\begin{aligned}\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta(x) = &\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}\,\zeta^s(u)](x)\\ &+ (1+ix)\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)](x)\\ &+ \mathcal{F}^{-1}[(\varphi_\varepsilon^{-1})'(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)](x).\end{aligned} \tag{8.17}$$

While $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta$ may exist only in distributional sense in general, it is defined rigorously through the right–hand side of the above display for $\zeta \in Z^{\gamma_s,\gamma_c}$. Considering $\zeta * K_h$ instead of ζ, we estimate separately all three terms in the following.

The continuity and linearity of the Fourier multiplier $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)]$, which was shown in Lemma 8.8(i), yield for the first term in (8.17)

$$\begin{aligned}\|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}\,\zeta^s(u)\,\mathcal{F}\,K_h(u)]\|_{H^\delta} &= \|\,\mathcal{F}^{-1}\,[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}[\zeta^s * K_h]]\,\|_{B^\delta_{2,2}}\\ &\lesssim \|\zeta^s * K_h\|_{B^{\beta+\delta}_{2,2}} \lesssim \|\zeta^s\|_{H^{\beta+\delta}},\end{aligned}$$

where the last inequality holds by $\|\,\mathcal{F}\,K_h\|_\infty \leqslant \|K\|_{L^1}$. Using the boundedness of f_Y and the continuous Sobolev embedding $H^{\delta/4}(\mathbb{R}) \subseteq L^{2+\delta}(\mathbb{R})$ by (8.31), (8.34) and (8.33), we obtain

$$\begin{aligned}&\|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}\,\zeta^s(u)\,\mathcal{F}\,K_h(u)]\|_{L^{2+\delta}(\mathbb{P})}\\ &\quad\lesssim \|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}\,\zeta^s(u)\,\mathcal{F}\,K_h(u)]\|_{L^{2+\delta}}\\ &\quad\lesssim \|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}\,\zeta^s(u)\,\mathcal{F}\,K_h(u)]\|_{H^\delta}\\ &\quad\lesssim \|\zeta^s\|_{H^{\beta+\delta}}\end{aligned} \tag{8.18}$$

To estimate the second term in (8.17), we use the Cauchy–Schwarz inequality and Assumption 8.4(ii):

$$\begin{aligned}&\|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}[\tfrac{1}{ix+1}\zeta^c(x)](u)\,\mathcal{F}\,K_h(u)]\|_\infty\\ &\quad\leqslant \|\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}[\tfrac{1}{ix+1}\zeta^c]\,\mathcal{F}\,K_h(u)\|_{L^1}\\ &\quad\lesssim \|\langle u\rangle^{-1/2-\beta-\delta}\varphi_\varepsilon^{-1}(-u)\|_{L^2}\|\langle u\rangle^{1/2+\beta+\delta}\,\mathcal{F}[\tfrac{1}{ix+1}\zeta^c(x)]\|_{L^2}\\ &\quad\lesssim \|\tfrac{1}{ix+1}\zeta^c(x)\|_{H^{1/2+\beta+\delta}}.\end{aligned}$$

Thus $\int(1+x^2)^{(2+\delta)/2}f_Y(x)\,\mathrm{d}x < \infty$ from (8.15) yields

$$\begin{aligned}&\|(1+ix)\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta^c(y)](u)\,\mathcal{F}\,K_h(u)](x)\|_{L^{2+\delta}(\mathbb{P})}\\ &\quad\lesssim \|\tfrac{1}{ix+1}\zeta^c(x)\|_{H^{1/2+\beta+\delta}}.\end{aligned} \tag{8.19}$$

The last term in the decomposition (8.17) can be estimated similarly using the Cauchy–

Schwarz inequality and Assumption 8.4(ii) for $(\varphi^{-1})'$

$$\begin{aligned}
&\| \mathcal{F}^{-1}[(\varphi_\varepsilon^{-1})'(-u)\,\mathcal{F}[\tfrac{1}{ix+1}\zeta^c(x)](u)\,\mathcal{F}K_h(u)]\|_{L^{2+\delta}(\mathbb{P})}\\
&\quad\lesssim \|(\varphi_\varepsilon^{-1})'(-u)\,\mathcal{F}[\tfrac{1}{ix+1}\zeta^c(x)](u)\|_{L^1}\\
&\quad\leqslant \|\langle u\rangle^{1/2-\beta-\delta}(\varphi_\varepsilon^{-1})'\|_{L^2}\|\langle u\rangle^{-1/2+\beta+\delta}\,\mathcal{F}^{-1}[\tfrac{1}{ix+1}\zeta^c(x)](u)\|_{L^2}\\
&\quad\lesssim \|\tfrac{1}{ix+1}\zeta^c(x)\|_{H^{-1/2+\beta+\delta}}.
\end{aligned}\tag{8.20}$$

Combining (8.18), (8.19) and (8.20), we obtain

$$\sup_{h\in(0,1)}\| \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}K_h]*\zeta(x)\|_{L^{2+\delta}(\mathbb{P})}\lesssim \|\zeta\|_{Z^{\beta+\delta,1/2+\beta+\delta}},\tag{8.21}$$

which is finite for δ small enough satisfying $\beta+\delta\leqslant\gamma_s$ and $1/2+\beta+\delta\leqslant\gamma_c$. Since $\mathcal{F}K_h$ converges pointwise to one and $|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}K_h]*\zeta(x)|^2$ is uniformly integrable by the bound of the $(2+\delta)$th moments, the variance converges to

$$\int\Big|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)]*\zeta(x)\Big|^2\,\mathbb{P}(\,\mathrm{d}x).$$

8.6 Tightness

Motivated by the representation (8.13) of the stochastic error, we introduce the empirical process

$$\nu_n(t):=\sqrt{n}\int\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}K_h]*\zeta_t(x)(\mathbb{P}_n-\mathbb{P})(\,\mathrm{d}x),\quad t\in\mathbb{R}\,.\tag{8.22}$$

In order to show tightness of the empirical process, we first show some properties of the class of translations $\mathcal{H}:=\{\zeta_t|t\in\mathbb{R}\}$ for $\zeta\in Z^{\gamma_s,\gamma_c}$.

Lemma 8.9. *For $\zeta\in Z^{\gamma_s,\gamma_c}$ the following is satisfied:*

(i) The decomposition $\zeta_t=\zeta_t^c+\zeta_t^s$ satisfies the conditions in the definition of Z^{γ_s,γ_c} with a_t. We have $\sup_{t\in\mathbb{R}}\|\zeta_t\|_{Z^{\gamma_s,\gamma_c}}<\infty$.

(ii) For any $\eta\in(0,\gamma_s)$, there exists $\tau>0$ such that $\|\zeta_t-\zeta_s\|_{Z^{\gamma_s-\eta,\gamma_c-\eta}}\lesssim|t-s|^\tau$ holds for all $s,t\in\mathbb{R}$ with $|t-s|\leqslant 1$.

Proof.

(i) Since $\|\zeta_t^s\|_{H^{\gamma_s}}^2=\int\langle u\rangle^{2\gamma_s}|e^{itu}\,\mathcal{F}\zeta^s(u)|^2\,\mathrm{d}u=\|\zeta^s\|_{H^{\gamma_s}}^2$, both claims hold for the singular part. Applying the pointwise multiplier property of Besov spaces (8.35) as well as the Besov embeddings (8.34) and (8.32), we obtain for some $M>\gamma_c$ and $a\in C^\infty(\mathbb{R})$ as in definition (8.7)

$$\begin{aligned}
\|\langle x\rangle^\tau(\zeta_t^c(x)-a_t(x))\|_{H^{\gamma_c}}&\lesssim\|\tfrac{\langle x\rangle^\tau}{\langle x-t\rangle^\tau}\|_{C^M}\|\langle x-t\rangle^\tau(\zeta_t^c(x)-a_t(x))\|_{H^{\gamma_c}}\\
&=\|\tfrac{\langle x\rangle^\tau}{\langle x-t\rangle^\tau}\|_{C^M}\|\langle x\rangle^\tau(\zeta^c(x)-a(x))\|_{H^{\gamma_c}},
\end{aligned}$$

which is finite for all $t \in \mathbb{R}$ since $\langle x \rangle^{\tau} \langle x-t \rangle^{-\tau} \in C^M(\mathbb{R})$. For the second claim we estimate similarly

$$\begin{aligned}\sup_{t\in\mathbb{R}} \|\tfrac{1}{ix+1}\zeta_t^c(x)\|_{H^{\gamma_c}} &\lesssim \sup_{t\in\mathbb{R}} \|\tfrac{a_t(x)}{ix+1}\|_{H^{\gamma_c}} + \|\tfrac{1}{ix+1}\|_{C^M} \sup_{t\in\mathbb{R}} \|\zeta_t^c - a_t\|_{H^{\gamma_c}} \\ &\lesssim \|\tfrac{1}{ix+1}\|_{H^{\gamma_c}} \|a\|_{C^M} + \|\tfrac{1}{ix+1}\|_{C^M} \|\zeta^c - a\|_{H^{\gamma_c}} < \infty.\end{aligned}$$

(ii) For the singular part note that

$$\begin{aligned}&\|\zeta_t^s - \zeta_s^s\|_{H^{\gamma_s-\eta}} \\ &\leqslant \|\langle u\rangle^{\gamma_s} \mathcal{F}\zeta^s(u)\|_{L^2} \|\langle u\rangle^{-\eta}(1-e^{i(t-s)u})\|_\infty \\ &\lesssim \|\langle u\rangle^{-\eta}\|_{L^\infty(\mathbb{R}\setminus(-|t-s|^{-1/2},|t-s|^{-1/2}))} \\ &\qquad \vee \|(1-e^{i(t-s)u})\|_{L^\infty((-|t-s|^{-1/2},|t-s|^{-1/2}))} \\ &\lesssim |t-s|^{\eta/2} \vee |t-s|^{1/2}.\end{aligned}$$

For ζ^c we have

$$\begin{aligned}\Big\|\tfrac{1}{ix+1}(\zeta_t^c(x) - \zeta_s^c(x))\Big\|_{H^{\gamma_c-\eta}} &\lesssim \Big\|\tfrac{1}{ix+1}\zeta_t^c(x) - \big(\tfrac{1}{ix+1}\zeta_t^c(x)\big) * \delta_{s-t}\Big\|_{H^{\gamma_c-\eta}} \\ &\quad + \Big\|\tfrac{1}{i(x-s+t)+1}\zeta_s^c(x) - \tfrac{1}{ix+1}\zeta_s^c(x)\Big\|_{H^{\gamma_c-\eta}}.\end{aligned}$$

The first term can be treated analogously to ζ^s. Using some integer $M \in \mathbb{N}$ strictly larger than γ_c, the second term can be estimated by

$$\begin{aligned}&\Big\|\tfrac{1}{i(x-s+t)+1}\zeta_s^c(x) - \tfrac{1}{ix+1}\zeta_s^c(x)\Big\|_{H^{\gamma_c-\eta}} \\ &\quad\lesssim |t-s| \Big\|\tfrac{1}{i(x-s+t)+1}\tfrac{1}{ix+1}\zeta_s^c(x)\Big\|_{H^{\gamma_c-\eta}} \\ &\quad\lesssim |t-s| \Big\|\tfrac{1}{i(x-s+t)+1}\Big\|_{C^M} \Big\|\tfrac{1}{ix+1}\zeta_s^c(x)\Big\|_{H^{\gamma_c-\eta}} \\ &\quad\lesssim |t-s|,\end{aligned}$$

where we used again pointwise multiplier (8.35), embedding properties of Besov spaces (8.34) and (8.32) as well as (i). □

8.6.1 Pregaussian limit process

Let $\mathbb{G}$ be the stochastic process from Theorem 8.5. It induces the intrinsic covariance metric $d(s,t) := \mathbb{E}[(\mathbb{G}_s - \mathbb{G}_t)^2]^{1/2}$.

Theorem 8.10. *There exists a version of $\mathbb{G}$ with uniformly d–continuous sample paths almost surely and with $\sup_{t\in\mathbb{R}} |\mathbb{G}_t| < \infty$ almost surely.*

The proof of the theorem shows in addition that $\mathbb{R}$ is totally bounded with respect to d. The boundedness of the sample paths follows from the totally bounded index set and the uniform continuity. Further we conclude that $\mathcal{G}$ defined in (8.8) is $\mathbb{P}$–pregaussian by van der Vaart and Wellner (1996, p. 89). Thus $\mathbb{G}$ is a tight Borel random variable in $\ell^\infty(\mathbb{R})$ and the law of $\mathbb{G}$ is uniquely defined through the covariance structure and the sample path properties in the theorem (van der Vaart and Wellner, 1996, Lem. 1.5.3).

Proof. To show that the class is pregaussian, it suffices to verify polynomial covering numbers. To that end, we deduce that

$$d(s,t) = \big(\|g_t - g_s\|^2_{L^2(\mathbb{P})} - \langle \zeta_t - \zeta_s, f_X\rangle^2\big)^{1/2} \leqslant \|g_t - g_s\|_{L^2(\mathbb{P})} \tag{8.23}$$

decreases polynomial for $|t-s| \to 0$, for $\max(s,t) \to -\infty$ and for $\min(s,t) \to \infty$. Using the same estimates which show the moment bound (8.21) but replacing $\mathcal{F} K_h = 1$, we obtain

$$\| \mathcal{F}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta\|_{L^2(\mathbb{P})} \lesssim \|\zeta\|_{Z^{\beta+\delta,1/2+\beta+\delta}} \tag{8.24}$$

and thus by choosing δ and η small enough Lemma 8.9 yields the bound $d(s,t) \lesssim \|\zeta_t - \zeta_s\|_{Z^{\beta+\delta,1/2+\beta+\delta}} \lesssim |t-s|^\tau$. We now turn to the estimation of the tails. We will only consider the case $s,t \geqslant N$ since the case $s,t \leqslant N$ can be treated in the same way. Without loss of generality, let $s < t$.

For the smooth component of ζ we have to show that $\big\|\frac{1}{ix+1}(\zeta_t^c(x) - \zeta_s^c(x))\big\|_{H^{\gamma_c}}$ with $t,s \geqslant N$ decays polynomially in N. It suffices to prove $\big\|\frac{1}{ix+1}(\zeta_t^c - a_t)(x)\big\|_{H^{\gamma_c}}$ and $\big\|\frac{1}{ix+1}(a_t - a_s)(x)\big\|_{H^{\gamma_c}}$ with $a \in C^\infty(\mathbb{R})$ from definition (8.7) of Z^{γ_s,γ_c} both decay polynomially in N. Let $M > \gamma_c$ and $\psi \in C^M(\mathbb{R})$ with $\psi(x) = 1$ for $x \in \mathbb{R}\setminus[-\frac{1}{2},\frac{1}{2}]$ and $\psi(x) = 0$ for $x \in [-\frac{1}{4},\frac{1}{4}]$. The pointwise multiplier property (8.35) yields

$$\begin{aligned}
&\big\|\tfrac{1}{ix+1}(\zeta_t^c - a_t)(x)\big\|_{H^{\gamma_c}}\\
&\quad= \big\|\big(\psi(x/N) + (1-\psi(x/N))\big)\tfrac{1}{ix+it+1}(\zeta^c - a)(x)\big\|_{H^{\gamma_c}}\\
&\quad\lesssim \|\tfrac{1}{ix+it+1}\|_{C^M}\|\psi(x/N)(\zeta^c - a)(x)\|_{H^{\gamma_c}} + \|\tfrac{1-\psi(x/N)}{ix+it+1}\|_{C^M}\|\zeta^c - a\|_{H^{\gamma_c}}\\
&\quad\lesssim \|\langle x\rangle^{-\tau}\psi(x/N)\|_{C^M}\|\langle x\rangle^{\tau}(\zeta^c - a)(x)\|_{H^{\gamma_c}} + N^{-1}\|\zeta^c - a\|_{H^{\gamma_c}}\\
&\quad\lesssim N^{-(\tau\wedge 1)}
\end{aligned}$$

and for N large enough such that $\operatorname{supp}(a') \subseteq [-N/2, N/2]$ we obtain

$$\begin{aligned}
&\big\|\tfrac{1}{ix+1}(a_t - a_s)(x)\big\|_{H^{\gamma_c}}\\
&\quad= \big\|\tfrac{\psi(x/N)}{ix+1}(a_t - a_s)(x)\big\|_{H^{\gamma_c}} \lesssim \big\|\tfrac{\psi(x/N)}{ix+1}\big\|_{H^{\gamma_c}}\big\|(a_t - a_s)(x)\big\|_{C^M}\\
&\quad\lesssim \big\|(ix+1)^{-3/4}\big\|_{H^{\gamma_c}}\big\|\psi(x/N)(ix+1)^{-1/4}\big\|_{C^M} \lesssim N^{-1/4}.
\end{aligned}$$

To bound the singular part it suffices to show that

$$\Big\|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta_t^s\Big\|_{L^2(\mathbb{P})}, \qquad t \geqslant N,$$

decays polynomially in N. To this end, we split the integral domain into

$$\left\| \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta_t^s \right\|_{L^2(\mathbb{P})}^2 = \int_{-\infty}^{-N/2} |\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,\zeta^s](x)|^2 f_Y(x+t)\,\mathrm{d}x + \int_{-N/2}^{\infty} |\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,\zeta^s](x)|^2 f_Y(x+t)\,\mathrm{d}x. \quad (8.25)$$

To estimate the first term, we use the following auxiliary calculations

$$\begin{aligned} &ix\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,\zeta^s](x) \\ &\quad = -\,\mathcal{F}^{-1}[(\varphi_\varepsilon^{-1})'(-\bullet)\,\mathcal{F}\,\zeta^s](x) + \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}[iy\zeta^s(y)]](x) \end{aligned}$$

and with an integer $M \in \mathbb{N}$ strictly larger than γ_s and a function $\chi \in C^M(\mathbb{R})$ which is equal to one on $\operatorname{supp}(\zeta^s)$ and has compact support

$$\begin{aligned} \|y\zeta^s(y)\|_{H^{\gamma_s}} = \|y\chi(y)\zeta^s(y)\|_{H^{\gamma_s}} &\lesssim \|y\chi(y)\|_{B^{\gamma_s}_{\infty,2}} \|\zeta^s(y)\|_{H^{\gamma_s}} \\ &\lesssim \|y\chi(y)\|_{C^M} < \infty, \end{aligned}$$

where we used the pointwise multiplier property (8.35) of Besov spaces as well as the Besov embeddings (8.34) and (8.32). Thus $ix\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,\zeta^s](x) \in L^2(\mathbb{R})$. Applying this and the boundedness of f_Y to the first term in (8.25) yields

$$\begin{aligned} &\int_{-\infty}^{-N/2} |\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,\zeta^s](x)|^2 f_Y(x+t)\,\mathrm{d}x \lesssim \int_{-\infty}^{-N/2} |\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,\zeta^s](x)|^2\,\mathrm{d}x \\ &\quad \leqslant 4N^{-2} \int_{-\infty}^{-N/2} |x\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,\zeta^s](x)|^2\,\mathrm{d}x \lesssim N^{-2}. \end{aligned}$$

Using Hölders's inequality and the boundedness of f_Y, we estimate the second term in (8.25) by

$$\begin{aligned} &\|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,\zeta^s](x)\|_{L^{2+\delta}}^2 \left(\int_{-N/2}^{\infty} |f_Y(x+t)|^{(2+\delta)/\delta}\,\mathrm{d}x \right)^{\delta/(2+\delta)} \\ &\lesssim \|\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,\zeta^s](x)\|_{L^{2+\delta}}^2 \left(\int_{N/2}^{\infty} f_Y(x)\,\mathrm{d}x \right)^{\delta/(2+\delta)}. \end{aligned}$$

While the first factor is finite according to our bound (8.18), which also holds when $\mathcal{F}\,K_h$ is omitted, the second one is of order $N^{-\delta}$ due to the finite $(2+\delta)$th moment of $\mathbb{P}$. Therefore, the second term in (8.25) decays polynomially. □

8.6.2 Uniform central limit theorem

We recall the definition of the empirical process ν_n in (8.22).

Theorem 8.11. *Grant Assumptions 8.1 and 8.4. Let*

$$(\nu_n(t_1),\ldots,\nu_n(t_k)) \xrightarrow{\mathcal{L}} (\mathbb{G}_{t_1},\ldots,\mathbb{G}_{t_k})$$

for all $t_1,\ldots,t_k \in \mathbb{R}$ *and for all* $k \in \mathbb{N}$. *If either* $\gamma_s \leqslant \beta + 1/2$ *and* $h_n^\rho n^{1/4} \to \infty$ *as* $n \to \infty$ *for some* $\rho > \beta - \gamma_s + 1/2$ *or if* $\gamma_s > \beta + 1/2$, *then*

$$\nu_n \xrightarrow{\mathcal{L}} \mathbb{G} \quad \text{in } \ell^\infty(\mathbb{R}).$$

Proof. We split the empirical process ν_n into three parts

$$\nu_n = \sqrt{n}\int (T_1(x) + T_2(x) + T_3(x))(\mathbb{P}_n - \mathbb{P})(\,\mathrm{d}x),$$

where T_1, T_2 and T_3 correspond to the three terms in decomposition (8.17) and are given by (8.26), (8.27) and (8.28) below. For the first term

$$T_1(x) = \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}\,\zeta_t^s(u)\,\mathcal{F}\,K_h(u)](x) \tag{8.26}$$

we distinguish the two cases $\gamma_s > \beta + 1/2$ and $\gamma_s \leqslant \beta + 1/2$. In the first case, we will show that T_1 varies in a fixed Donsker class. In the second case, the process indexed by T_1 is critical, this is where smoothed empirical processes and the condition on the bandwidth are needed. Tightness of T_1 in this case will be shown in Section 8.6.3. We will further show that the second term T_2 and the third term T_3 are both varying in fixed Donsker classes for all $\gamma_s > \beta$. In particular the three processes indexed by T_1, T_2 and T_3, respectively, are tight. Applying the equicontinuity characterization of tightness (van der Vaart and Wellner, 1996, Thm. 1.5.7) with the maximum of the semimetrics yields that ν_n is tight. Since we have assumed convergence of the finite dimensional distribution, the convergence of ν_n in distribution follows (van der Vaart and Wellner, 1996, Thm. 1.5.4).

Here we consider only the first case of $\gamma_s > \beta + 1/2$. We recall that ζ_t^s is contained in $H^{\gamma_s}(\mathbb{R})$. By the Fourier multiplier property of the deconvolution operator in Lemma 8.8(i) and by $\sup_{h>0,u} |\,\mathcal{F}\,K_h(u)| \leqslant \|K\|_{L^1} < \infty$ the functions T_1 are contained in a bounded set of $H^{1/2+\eta}(\mathbb{R})$ for some $\eta > 0$ small enough. We apply (Nickl and Pötscher, 2007, Prop. 1) with $p = q = 2$ and $s = 1/2 + \eta$ and conclude that T_1 varies in a universal Donsker class.

The second term is of the form

$$T_2(x) = (1 + ix)\,\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta_t^c(y)](u)\,\mathcal{F}\,K_h(u)](x). \tag{8.27}$$

By Assumption 8.4(ii) we have $\varphi_\varepsilon^{-1}(u) \lesssim \langle u\rangle^\beta$. For some $\eta > 0$ sufficiently small, the functions $\frac{1}{iy+1}\zeta_t^c(y)$, $t \in \mathbb{R}$, are contained in a bounded set of $H^{\beta+\eta+1/2}(\mathbb{R})$ by Lemma 8.9. We obtain that the functions $T_2(x)/(1+ix)$ are contained in a bounded subset of $H^{1/2+\eta}(\mathbb{R})$. Corollary 5 in (Nickl and Pötscher, 2007) yields with $p = q = 2$, $\beta = -1$, $s = 1/2 + \eta$ and $\gamma = \eta$ that T_2 is contained in a fixed $\mathbb{P}$–Donsker class.

Similarly, we treat the third term

$$T_3(x) = \mathcal{F}^{-1}[(\varphi_\varepsilon^{-1})'(-u)\,\mathcal{F}[\tfrac{1}{iy+1}\zeta_t^c(y)](u)\,\mathcal{F}\,K_h(u)](x). \tag{8.28}$$

By Assumption 8.4(ii) we have $(\varphi_\varepsilon^{-1})' \lesssim \langle u\rangle^{\beta-1}$. As above we conclude that the functions T_3 are contained in a bounded set of $H^{\eta+3/2}(\mathbb{R})$. By (Nickl and Pötscher, 2007, Prop. 1) with $p = q = 2$ and $s = \eta + 3/2$ the term T_3 varies in a universal Donsker class. □

8.6.3 The critical term

In this section, we treat the first term T_1 in the case $\gamma_s \leqslant \beta + 1/2$. We define

$$q_t := \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}\,\zeta_t^s(u)]. \tag{8.29}$$

For simplicity in point (e) below it will be convenient to work with functions K_h of bounded support. Thus we fix $\xi > 0$ and define the truncated kernel

$$K_h^{(0)} := K_h \mathbf{1}_{[-\xi,\xi]}.$$

By the Assumption (8.6) on the decay of K we have $\sup_{h>0} \|K_h - K_h^{(0)}\|_{BV} < \infty$. We conclude $\mathcal{F}(K_h - K_h^{(0)})(u) \lesssim (1+|u|)^{-1}$ with a constant independent of $h > 0$. By Assumption 8.4(ii) we have $|\varphi_\varepsilon^{-1}(u)| \lesssim (1+|u|)^\beta$. The functions $\zeta_t^s(u)$, $t \in \mathbb{R}$, are contained in a bounded set of $H^{\gamma_s}(\mathbb{R})$. Consequently, T_1 with $K_h - K_h^{(0)}$ instead of K_h is contained in a bounded set of $H^{\gamma_s-\beta+1}(\mathbb{R})$. With the same argument as used for T_3, we see that this term is contained in a universal Donsker class because $\gamma_s - \beta + 1 > 1$ by assumption. So it remains to consider T_1 with the truncated kernel $K_h^{(0)}$.

In order to show tightness of the process indexed by T_1 with the truncated kernel $K_h^{(0)}$, we check the assumptions of Theorem 3 by Giné and Nickl (2008) in the version of Nickl and Reiß (2012, Thm. 12) for the class $Q = \{q_t | t \in \mathbb{R}\}$ and for $\mu_n(\,\mathrm{d}x) := K_{h_n}^{(0)}(x)\,\mathrm{d}x$, where $q_t(x)$ was defined in (8.29). By Section 8.6.1 we know that the class $\mathcal{G}$ is $\mathbb{P}$–pregaussian. From the proof also follows that Q is $\mathbb{P}$–pregaussian since this is just the case $\zeta^c = 0$.

We write

$$Q'_\tau := \{r - q | r, q \in Q, \|r - q\|_{L^2(\mathbb{P})} \leqslant \tau\}.$$

Let $\rho > \beta - \gamma_s + 1/2 \geqslant 0$ be such that $h_n^\rho n^{1/4} \to \infty$. We fix some $\rho' \in (\beta - \gamma_s + 1/2, \rho \wedge 1)$ and obtain $h_n^{\rho'} \log(n)^{-1/2} n^{1/4} \to \infty$. We need to verify the following conditions.

(a) We will show that the functions in $\widetilde{Q}_n := \{q_t * \mu_n | t \in \mathbb{R}\}$ are bounded by $M_n := C h_n^{-\rho'}$ for some constant $C > 0$. Since q_t is only a translation of q_0 it suffices to consider q_0. By the definition of Z^{γ_s,γ_c} in (8.7), by Lemma 8.8(i) and by the Besov embedding (8.34) we have

$$q_0 = \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)\,\mathcal{F}\,\zeta^s(u)] \in B_{2,2}^{\gamma_s-\beta}(\mathbb{R}) \subseteq B_{2,\infty}^{1/2-\rho'}(\mathbb{R}).$$

By our assumptions on the kernel (8.6) it follows that K' is integrable and thus that K is of bounded variation. Next, we apply continuous embeddings for Besov spaces (8.31) and (8.33), (8.36) as well as the estimate for $\|K_{h_n}\|_{B_{1,1}^{\rho'}}$ in Giné and Nickl (2008, p. 384), which also applies to truncated kernels, and obtain

$$\|q_0 * K_{h_n}^{(0)}\|_\infty \lesssim \|q_0 * K_{h_n}^{(0)}\|_{B_{\infty,1}^0} \lesssim \|q_0 * K_{h_n}^{(0)}\|_{B_{2,1}^{1/2}} \lesssim \|K_{h_n}^{(0)}\|_{B_{1,1}^{\rho'}} \lesssim h_n^{-\rho'}. \quad (8.30)$$

(b) For $r \in Q'_\tau$ holds $\|r * K_h^{(0)}\|_{L^2(\mathbb{P})} \leqslant \|r * K_h^{(0)} - r\|_{L^2(\mathbb{P})} + \tau$. Thus it suffices to show that $\|q * K_h^{(0)} - q\|_{L^2(\mathbb{P})} \to 0$ uniformly over $q \in Q$. We estimate

$$\|q_t * K_h^{(0)} - q_t\|_{L^2(\mathbb{P})} \lesssim \|\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,\zeta^s(\mathcal{F}\,K_h^{(0)} - 1)\|_{L^2}.$$

$\varphi_\varepsilon^{-1}(-\bullet)\,\mathcal{F}\,\zeta^s$ is an L^2–function and $\mathcal{F}\,K_h^{(0)}$ is uniformly bounded and converges to one as $h \to 0$. By dominated convergence the integral converges to zero.

(c) The estimates in (a) can be used to see that the classes $\widetilde{Q}_n$ have polynomial $L^2(\mathbb{Q})$–covering numbers, uniformly in all probability measures $\mathbb{Q}$ and uniformly in n. The function $q_0 * K_{h_n}^{(0)}$ is the convolution of two L^2–functions and thus continuous. The estimate (8.30) and embedding (8.37) yield that $q_0 * K_{h_n}^{(0)}$ is of finite 2–variation. We argue as in Lemma 1 by Giné and Nickl (2009). As a function of bounded 2–variation $q_0 * K_{h_n}^{(0)}$ can be written as a composition $g_n \circ f_n$ of a nondecreasing function f_n and a function g_n, which satisfies a Hölder condition $|g_n(u) - g_n(v)| \leqslant |u - v|^{1/2}$, see, for example, (Dudley, 1992, p. 1971). More precisely, we can take $f_n(x)$ to be the 2–variation of $q_0 * K_{h_n}^{(0)}$ up to x and the envelopes of f_n to be multiples of $M_n^2 = C^2 h_n^{-2\rho'}$. The set F_n of all translates of the nondecreasing function f_n has VC–index 2 and thus polynomial $L^1(\mathbb{Q})$–covering numbers (de la Peña and Giné, 1999, Thm. 5.1.15). Since each ε^2–covering of translates of f_n for $L^1(\mathbb{Q})$ induces an ε–covering of translates of $g_n \circ f_n$ for $L^2(\mathbb{Q})$, we can estimate the covering numbers by

$$N(\widetilde{Q}_n, L^2(\mathbb{Q}), \varepsilon) \leqslant N(F_n, L^1(\mathbb{Q}), \varepsilon^2) \lesssim (M_n/\varepsilon)^4,$$

with constants independent of n and $\mathbb{Q}$. The conditions for inequality (22) by Giné and Nickl (2008) are fulfilled, where the envelopes are $M_n = C h_n^{-\rho'}$ and

$H_n(\eta) = H(\eta) = C_1 \log(\eta) + C_0$ with $C_0, C_1 > 0$. Consequently

$$\mathbb{E}^* \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \varepsilon_j f(X_j) \right\|_{(\widetilde{Q}_n)'_{n^{-1/4}}} \lesssim \max\left(\frac{\sqrt{\log(n)}}{n^{1/4}}, \frac{h_n^{-\rho'}}{\sqrt{n}} \log(n) \right) \to 0$$

as $n \to \infty$.

(d) We apply Lemma 1 of Giné and Nickl (2008) to show that

$$\cup_{n\geqslant 1} \widetilde{Q}_n = \bigcup_{n\geqslant 1} \left\{ x \mapsto \int_{\mathbb{R}} q_t(x-y) K_{h_n}^{(0)}(y)\,\mathrm{d}y \,\middle|\, t \in \mathbb{R} \right\}$$

is in the $L^2(\mathbb{P})$–closure of $\|K\|_{L^1}$–times the symmetric convex hull of the pregaussian class Q. The condition $q_t(\bullet - y) \in L^2(\mathbb{P})$ is satisfied for all $y \in \mathbb{R}$ since $q_t \in L^2(\mathbb{R})$ and f_Y is bounded. $q_t(x - \bullet) \in L^1(|\mu_n|)$ is fulfilled owing to $K_{h_n}^{(0)}, q_t \in L^2(\mathbb{R})$. The third condition that $y \mapsto \|q_t(\bullet - y)\|_{L^2(\mathbb{P})}$ is in $L^1(|\mu_n|)$ holds likewise since f_Y is bounded and $K_{h_n}^{(0)} \in L^1(\mathbb{R})$.

(e) The $L^2(\mathbb{P})$–distance of two functions in $\widetilde{Q}_n$ can be estimated by

$$\begin{aligned}
&\mathbb{E}\left[(q_t * K_h^{(0)}(X) - q_s * K_h^{(0)}(X))^2 \right]^{1/2} \\
&= \left\| \int q_t(\bullet - u) K_h^{(0)}(u) - q_s(\bullet - u) K_h^{(0)}(u)\,\mathrm{d}u \right\|_{L^2(\mathbb{P})} \\
&\leqslant \int |K_h^{(0)}(u)| \|q_t(\bullet - u) - q_s(\bullet - u)\|_{L^2(\mathbb{P})}\,\mathrm{d}u \\
&\leqslant \|K_h^{(0)}\|_{L^1} \sup_{|u|\leqslant \xi} \|q_t(\bullet - u) - q_s(\bullet - u)\|_{L^2(\mathbb{P})} \\
&= \|K_h^{(0)}\|_{L^1} \sup_{|u|\leqslant \xi} \|q_{t+u} - q_{s+u}\|_{L^2(\mathbb{P})}.
\end{aligned}$$

As seen in the proof that Q is pregaussian, the covering numbers grow at most polynomially. We take N large enough such that $N \geqslant 2\xi$. Then $s, t > N$ implies $s + u, t + u > N/2$ and $s, t < -N$ implies $s + u, t + u < -N/2$. Since this is only a polynomial change in N, the growth of the covering numbers remains at most polynomial. This leads to the entropy bound $H(\widetilde{Q}_n, L^2(\mathbb{P}), \eta) \lesssim \log(\eta^{-1})$ for η small enough and independent of n. We define $\lambda_n(\eta) := \log(\eta^{-1})\eta^2$. The bound in the condition is of the order $\log(n)^{-1/2} n^{1/4}$. As seen before (a) this growth faster than $M_n = C h_n^{-\rho'}$.

8.7 Function spaces

Here we summarize definitions and properties of function spaces used in this chapter. Let us define the L^p–Sobolev space for $p \in (0,\infty)$ and $m \in \mathbb{N}$

$$W_p^m(\mathbb{R}) := \Big\{ f \in L^p(\mathbb{R}) \Big| \sum_{k=0}^{m} \|f_X^{(k)}\|_{L^p} < \infty \Big\}$$

In particular, $W_p^0(\mathbb{R}) = L^p(\mathbb{R})$. Due to the Hilbert space structure, the case $p = 2$ is crucial. It can be described equivalently with the notation $\langle u \rangle = (1+u^2)^{1/2}$ by, $\alpha \geqslant 0$,

$$H^\alpha(\mathbb{R}) := \Big\{ f \in L^2(\mathbb{R}) \Big| \|f\|_{H^\alpha}^2 := \int \langle u \rangle^{2\alpha} |\mathcal{F} f(u)|^2 \,\mathrm{d}u < \infty \Big\},$$

which we call Sobolev space, too. Obviously, $W_2^m(\mathbb{R}) = H^m(\mathbb{R})$. Also frequently used are the Hölder spaces. Denoting the space of all bounded, continuous functions with values in $\mathbb{R}$ as $C(\mathbb{R})$ we define, $\alpha \geqslant 0$,

$$C^\alpha(\mathbb{R}) := \Big\{ f \in C(\mathbb{R}) \Big| \|f\|_{C^\alpha} := \sum_{k=0}^{\lfloor \alpha \rfloor} \|f^{(l)}\|_\infty + \sup_{x \neq y} \frac{|f^{(\lfloor \alpha \rfloor)}(x) - f^{(\lfloor \alpha \rfloor)}(y)|}{|x-y|^{\alpha - \lfloor \alpha \rfloor}} < \infty \Big\},$$

where $\lfloor \alpha \rfloor$ denotes the largest integer smaller or equal to α. A unifying approach which contains all function spaces defined so far, is given by Besov spaces (Triebel, 2010, Sect. 2.3.1) which we will discuss in the sequel. Let $\mathscr{S}(\mathbb{R})$ be the Schwartz space of all rapidly decreasing infinitely differentiable functions with values in $\mathbb{C}$ and $\mathscr{S}'(\mathbb{R})$ its dual space, that is the space of all tempered distributions. Let $0 < \psi \in \mathscr{S}(\mathbb{R})$ with $\operatorname{supp} \psi \subseteq \{x | 1/2 \leqslant |x| \leqslant 2\}$ and $\psi(x) > 0$ if $\{x | 1/2 < |x| < 2\}$. Then define $\varphi_j(x) := \psi(2^{-j}x)(\sum_{k=-\infty}^{\infty} \psi(2^{-k}x))^{-1}, j = 1, 2, \ldots$, and $\varphi_0(x) := 1 - \sum_{j=1}^{\infty} \varphi_j(x)$ such that the sequence $\{\varphi_j\}_{j=0}^{\infty}$ is a smooth resolution of unity. In particular, $\mathcal{F}^{-1}[\varphi_j \mathcal{F} f]$ is an entire function for all $f \in \mathscr{S}'(\mathbb{R})$. For $s \in \mathbb{R}$ and $p, q \in (0,\infty]$ the Besov spaces are defined by

$$B_{p,q}^s := \Big\{ f \in \mathscr{S}'(\mathbb{R}) \Big| \|f\|_{B_{p,q}^s} := \Big(\sum_{j=0}^{\infty} 2^{sjq} \| \mathcal{F}^{-1}[\varphi_j \mathcal{F} f] \|_{L^p}^q \Big)^{1/q} < \infty \Big\}.$$

We omit the dependence of $\|\bullet\|_{B_{p,q}^s}$ on ψ since any function with the above properties defines an equivalent norm. Setting the Besov spaces in relation to the more elementary function spaces, we first note that the Schwartz functions $\mathscr{S}(\mathbb{R})$ are dense in every Besov space $B_{p,q}^s$ with $p, q < \infty$ and $H^\alpha(\mathbb{R}) = B_{2,2}^\alpha(\mathbb{R})$ as well as $C^\alpha(\mathbb{R}) = B_{\infty,\infty}^\alpha(\mathbb{R})$, where the latter holds only if α is not an integer (Triebel, 2010, Thms. 2.3.3 and 2.5.7). Frequently used are the following continuous embeddings, which can be found in (Triebel, 2010, Sect. 2.5.7, Thms. 2.3.2(1), 2.7.1): For $p \geqslant 1, m \in \mathbb{Z}$

$$B_{p,1}^m(\mathbb{R}) \subseteq W_p^m(\mathbb{R}) \subseteq B_{p,\infty}^m(\mathbb{R}) \quad \text{and} \quad B_{\infty,1}^0(\mathbb{R}) \subseteq L^\infty(\mathbb{R}) \subseteq B_{\infty,\infty}^0(\mathbb{R}) \tag{8.31}$$

and for $s \geqslant 0$

$$B^s_{\infty,1}(\mathbb{R}) \subseteq C^s(\mathbb{R}) \subseteq B^s_{\infty,\infty}(\mathbb{R}). \tag{8.32}$$

Furthermore, for $0 < p_0 \leqslant p_1 \leqslant \infty, q \geqslant 0$ and $-\infty < s_1 \leqslant s_0 < \infty$

$$B^{s_0}_{p_0,q}(\mathbb{R}) \subseteq B^{s_1}_{p_1,q}(\mathbb{R}) \quad \text{if} \quad s_0 - \frac{1}{p_0} \geqslant s_1 - \frac{1}{p_1} \tag{8.33}$$

and for $0 < p, q_0, q_1 \leqslant \infty$ and $-\infty < s_1 < s_0 < \infty$

$$B^{s_0}_{p,q_0}(\mathbb{R}) \subseteq B^{s_1}_{p,q_1}(\mathbb{R}). \tag{8.34}$$

Another important relation is the pointwise multiplier property of Besov spaces (Triebel, 2010, (24) on p. 143) that is

$$\|fg\|_{B^s_{p,q}} \lesssim \|f\|_{B^s_{\infty,q}} \|g\|_{B^s_{p,q}} \tag{8.35}$$

for $s > 0$, $1 \leqslant p \leqslant \infty$ and $0 < q \leqslant \infty$.

The Besov norm of a convolution can be bounded by Lemma 7 (i) in Qui (1981). Let $1 \leqslant p, q, r, s \leqslant \infty$, $-\infty < \alpha, \beta < \infty$, $0 \leqslant 1/u = 1/p + 1/r - 1 \leqslant 1$, $0 \leqslant 1/v = 1/q + 1/s \leqslant 1$. For $f \in B^\alpha_{p,q}(\mathbb{R})$ and $g \in B^\beta_{r,s}(\mathbb{R})$

$$\|f * g\|_{B^{\alpha+\beta}_{u,v}} \lesssim \|f\|_{B^\alpha_{p,q}} \|g\|_{B^\beta_{r,s}}. \tag{8.36}$$

Using for any function $f : \mathbb{R} \to \mathbb{R}$ and $h \in \mathbb{R}$ the difference operators $\Delta^1_h f(x) := f(x+h) - f(x)$ and $(\Delta^l_h f)(x) := \Delta^1_h(\Delta^{l-1}_h f)(x), l \in \mathbb{N}$, the Besov can be equivalently described by

$$\|f\|_{B^s_{pq}} \sim \|f\|_{L^p} + \|f\|_{\dot{B}^s_{pq}} \quad \text{with} \quad \|f\|_{\dot{B}^s_{pq}} := \Big(\int |h|^{-sq-1} \|\Delta^M_h f\|^q_{L^p} \,\mathrm{d}h\Big)^{1/q}$$

for $s > 0, p, q \geqslant 1$ and any integer $M > s$ (Triebel, 2010, Thm. 2.5.12). The space of all $f \in \mathscr{S}'(\mathbb{R})$ for which $\|f\|_{\dot{B}^s_{pq}}$ is finite is called homogeneous Besov space $\dot{B}^s_{pq}(\mathbb{R})$ (Triebel, 2010, Def. 5.1.3/2, Thm. 2.2.3/2) and thus $B^s_{pq} = L^p(\mathbb{R}) \cap \dot{B}^s_{pq}(\mathbb{R})$ for $s > 0, p, q \geqslant 1$. Of interest is the relation of homogeneous Besov spaces to functions of bounded p–variation. Let $\mathcal{BV}_p(\mathbb{R})$ denote the space of measurable functions $f : \mathbb{R} \to \mathbb{R}$ such that there is a function g which coincides with f almost everywhere and satisfies

$$\sup\Big\{ \sum_{i=1}^n |g(x_i) - g(x_{i-1})|^p \Big| -\infty < x_1 < \cdots < x_n < \infty, n \in \mathbb{N} \Big\} < \infty$$

and we define $BV_p(\mathbb{R})$ as the quotient set $\mathcal{BV}_p(\mathbb{R})$ modulo equality almost everywhere. Then,

$$\dot{B}^{1/p}_{p1}(\mathbb{R}) \subseteq BV_p(\mathbb{R}) \subseteq \dot{B}^{1/p}_{p,\infty}(\mathbb{R}), \qquad \text{for } p > 1 \tag{8.37}$$

by (Bourdaud et al., 2006, Thm. 5). For $p = 1$ holds by (Giné and Nickl, 2008, Lem. 8)

$$BV_1(\mathbb{R}) \cap L^1(\mathbb{R}) \subseteq B^1_{1,\infty}(\mathbb{R}). \tag{8.38}$$

9 Conclusion and outlook

The goal of the thesis was to extend the spectral calibration method by constructing confidence sets and to study a more fundamental uniform central limit theorem in the related deconvolution model. Both tasks have been accomplished. Along the way a number of aspects concerning the spectral calibration method have been investigated. The question on which conditions the estimators for observations in the Gaussian white noise model are well–defined has led to a problem on a particular Gaussian process. We have solved it by proving a general upper bound for hitting probabilities of Gaussian random fields in terms of Hausdorff dimensions. Joint asymptotic normality of the estimators has been proved. In view of honest confidence sets, we have extended the asymptotic normality results to be uniform in the underlying probability measure. Based on the asymptotic variance, asymptotic confidence sets have been constructed. We have adapted our results for practical purposes and our simulations show that this improves the performance of the method and of the confidence intervals. Finally, the procedures have been applied to option data of the German DAX index. We have seen that the problem of estimating the characteristic triplet of a Lévy process exhibits a deconvolution structure. Using the theory of smoothed empirical processes, we have proved a central limit theorem for estimators of linear functions in the deconvolution problem, which is uniform with respect to translations.

The calibration and the deconvolution problem have a similar structure. Let us take a closer look on the relation between our uniform central limit theorem in the deconvolution model and the estimation of characteristic triplets of Lévy processes. Since our result is for particular situations of the mildly ill–posed case, we are necessarily talking about the mildly ill–posed case for Lévy processes, too. In our set–up, this is the compound Poisson case with a possible drift. The connection to Lévy processes is given through the paper by Nickl and Reiß (2012), who estimate the generalized distribution function of Lévy measures based on low–frequency observations and prove a uniform central limit theorem for the estimators. In the compound Poisson case the deconvolution operator is given through the convolution with the finite signed measure

$$\mathcal{F}^{-1}[\varphi_T^{-1}(-\bullet)] = \sum_{k=0}^{\infty} \frac{e^{T\lambda}(-T)^k}{k!} \overline{\nu}^{*k} \quad \text{with } \overline{\nu}(A) := \nu(-A).$$

In the spirit of our analysis in the deconvolution model, there is an interplay between the smoothness of the functionals ζ_t and the mapping properties of the deconvolution operator. With appropriate smoothness assumptions on ν it should be possible to extend the results on Lévy processes from estimating the distribution function, $\zeta_t = \mathbf{1}_{(-\infty,t]}$, to estimating other functionals including the density, $\zeta_t = \delta_t$.

We would like to conclude by discussing interesting fields for further research, which open up when looking on our results from a wider perspective.

- **Uniform results for the severely ill–posed case**
Since the deconvolution problem and the estimation of characteristic triplets have a similar structure, it stands to reason to look at the existing results for the two problems jointly rather than separately. It is striking that, in the severely ill–posed case, there seem to be no uniform results for either one of the problems. This holds for the estimation of the density and as well for the estimation of the distribution function. To the best of our knowledge even joint asymptotic normality has not been studied before our work. However, van Es and Uh (2005) have considered interval probabilities in the severely ill–posed case of the deconvolution problem, which may be expressed by differences of the distribution functions at the endpoints of the intervals. So effects of the joint asymptotic distribution are implicitly contained in their work since the difference of two estimators is considered. For the estimators of interval probabilities they have found that the asymptotic behavior depends on the sequence of cut–off values. We find a similar dependence when looking at the joint asymptotic distribution of pointwise estimators for the jump density in the severely ill–posed case. This phenomenon completely vanishes when we consider only the asymptotic distribution of an individual estimator instead of the joint asymptotic distribution of the estimators. Our theorems for the severely ill–posed case yield an understanding how the asymptotic behavior of the estimators depends on the sequence of the cut–off values. It is an interesting question whether there are uniform results with a limit process in the usual form and, if not, how an alternative formulation of a uniform result would look like.

- **Lower bounds in the Lévy set–up**
Belomestny and Reiß (2006a) proved that the spectral calibration method is optimal in the minimax sense. Our analysis shows the precise structure of the asymptotic variance, in particular, how the noise level enters into the asymptotic variance. A natural follow–up question is to ask for lower bounds of the asymptotic variance. We have seen in the mildly ill–posed case that the asymptotic variance depends locally on the noise level. This suggests that it might be possible to define estimators that depend only on local properties of the option function $\mathcal{O}$. Indeed, that is true for drift γ and the intensity λ, which can be determined by a change point detection algorithm for jumps in the derivative of $\mathcal{O}$, see the observation after Proposition 2.1 in Belomestny and Reiß (2006a). This leads to the related question whether the exponentially weighted jump density $\mu(x)$ can be estimated from local properties of $\mathcal{O}$ at $x + T\gamma$.

- **Confidence bands in deconvolution**
The uniform central limit theorem is a key step for the construction of confidence bands. The theorem reduces the construction of confidence bands to analyzing the distribution of the limit process. More explicit knowledge on the distribution of the limit process would be very useful. The pointwise covariance of the limit

process can be estimated consistently. But feasible confidence bands call for further understanding what effect the use of the empirical covariance has.

- **Uniformity with respect to the underlying probability measure**
  For the estimation of characteristic triplets we have shown uniform convergence in the underlying probability distribution. This uniformity with respect to the underlying probability distribution is rarely treated although it is the basis for honest confidence sets and thus for sound confidence statements. For Donsker classes there are conditions for the uniform convergence in the probability distribution, see Section 2.8 in van der Vaart and Wellner (1996). Not only in view of honest confidence bands in deconvolution, a similar criterion for smoothed empirical processes would be desirable.

- **Adaptivity**
  We have not touched the question of adaptivity here. In the nonparametric problem of density estimation Low (1997) proved that confidence intervals which are honest and adaptive do not exist. Our confidence intervals are honest but not adaptive. Even if we consider only the estimation of a single parameter in the Lévy set–up, for example, the drift, it is an open question whether in our situation honest and adaptive confidence intervals exist. But also from a more practical perspective, the data–driven choice of the cut–off value should be further investigated.

- **Pricing and hedging error**
  Studying the impact of the calibration error on pricing and hedging is of great relevance for the application of the method in practice. For derivative pricing, Cont (2006) investigates the influence of model uncertainty in a general framework. More specific results on exponential Lévy models would be of interest.

In summary, we have covered a wide range of topics in the realm of the calibration and the deconvolution problem reaching from problems in probability theory to the application to option data. While the initial questions have been answered, others open up and provide interesting links for further research.

Acknowledgements

First of all, I would like to thank my advisor Markus Reiß for the guidance I have received while working on this thesis. I am especially thankful for his trust in my abilities from the very beginning, for many insightful comments and fruitful discussions during the whole time as well as for sharing his fascination for mathematics with me. I am grateful to Richard Nickl for giving me the possibility of a research stay in Cambridge and for intensive supervision during my stay, which considerably contributed to the progress of the project I started there. I am thankful to Mathias Trabs for productive collaboration with many discussions. In addition, I thank Denis Belomestny for helpful comments. This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 “Economic Risk”. I am thankful to the Statistical Laboratory in Cambridge for the hospitality during my stay there. I thank my friends and my family for their support, especially I am thankful to my parents.

Bibliography

Aït-Sahalia, Y. and J. Jacod (2009). Estimating the degree of activity of jumps in high frequency data. *Ann. Statist. 37*(5A), 2202 – 2244.

Barndorff-Nielsen, O. E. (1998). Processes of normal inverse Gaussian type. *Finance Stoch. 2*(1), 41–68.

Bauer, F. and M. Reiß (2008). Regularization independent of the noise level: an analysis of quasi–optimality. *Inverse Problems 24*(5), 1–16.

Belomestny, D. (2010). Spectral estimation of the fractional order of a Lévy process. *Ann. Statist. 38*(1), 317–351.

Belomestny, D. (2011). Statistical inference for time–changed Lévy processes via composite characteristic function estimation. *Ann. Statist. 39*(4), 2205–2242.

Belomestny, D. and M. Reiß (2006a). Spectral calibration of exponential Lévy models. *Finance Stoch. 10*(4), 449–474.

Belomestny, D. and M. Reiß (2006b). Spectral calibration of exponential Lévy models [2]. Discussion Paper 2006-035, SFB 649, Humboldt–Universität zu Berlin, Germany. Available at http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2006-035.pdf.

Belomestny, D. and J. Schoenmakers (2011). A jump–diffusion Libor model and its robust calibration. *Quant. Finance 11*(4), 529–546.

Bickel, P. J. and Y. Ritov (2003). Nonparametric estimators which can be "plugged–in". *Ann. Statist. 31*(4), 1033–1053.

Bickel, P. J. and M. Rosenblatt (1973). On some global measures of the deviations of density function estimates. *Ann. Statist. 1*, 1071–1095.

Biermé, H., C. Lacaux, and Y. Xiao (2009). Hitting probabilities and the Hausdorff dimension of the inverse images of anisotropic Gaussian random fields. *Bull. Lond. Math. Soc. 41*(2), 253–273.

Bissantz, N., L. Dümbgen, H. Holzmann, and A. Munk (2007). Non–parametric confidence bands in deconvolution density estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol. 69*(3), 483–506.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy 81*(3), 637–654.

Bourdaud, G., M. Lanza de Cristoforis, and W. Sickel (2006). Superposition operators and functions of bounded p–variation. *Rev. Mat. Iberoam. 22*(2), 455–487.

Brown, L. D. and M. G. Low (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist. 24*(6), 2384–2398.

Butucea, C. and F. Comte (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli 15*(1), 69–98.

Carr, P., H. Geman, D. B. Madan, and M. Yor (2002). The fine structure of asset returns: An empirical investigation. *J. Bus. 75*(2), 305–332.

Carr, P. and D. B. Madan (1999). Option valuation using the fast Fourier transform. *J. Comput. Finance 2*, 61–73.

Chung, K. L. (1974). *A course in probability theory* (Second ed.). Academic Press, New York–London. Probability and Mathematical Statistics, Vol. 21.

Comte, F. and V. Genon-Catalot (2009). Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Process. Appl. 119*(12), 4088–4123.

Comte, F. and V. Genon-Catalot (2010). Nonparametric adaptive estimation for pure jump Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat. 46*(3), 595–617.

Comte, F. and V. Genon-Catalot (2011). Estimation for Lévy processes from high frequency data within a long time interval. *Ann. Statist. 39*(2), 803–837.

Cont, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Math. Finance 16*(3), 519–547.

Cont, R. and P. Tankov (2004a). *Financial modelling with jump processes.* Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL.

Cont, R. and P. Tankov (2004b). Non–parametric calibration of jump–diffusion option pricing models. *J. Comput. Finance 7*(3), 1–49.

Cont, R. and P. Tankov (2006). Retrieving Lévy processes from option prices: regularization of an ill–posed inverse problem. *SIAM J. Control Optim. 45*(1), 1–25 (electronic).

Dattner, I., A. Goldenshluger, and A. Juditsky (2011). On deconvolution of distribution functions. *Ann. Statist. 39*(5), 2477–2501.

de la Peña, V. H. and E. Giné (1999). *Decoupling: From Dependence to Independence.* Springer, New York.

Dette, H. and N. Neumeyer (2001). Nonparametric analysis of covariance. *Ann. Statist. 29*(5), 1361–1400.

Dudley, R. M. (1992). Fréchet differentiability, p–variation and uniform Donsker classes. *Ann. Probab. 20*(4), 1968–1982.

Dudley, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge University Press, Cambridge.

Eberlein, E., U. Keller, and K. Prause (1998). New insights into smile, mispricing and value at risk: The hyperbolic model. *J. Bus. 71*, 371–406.

Fan, J. (1991a). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A 53*(1), 97–110.

Fan, J. (1991b). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist. 19*(3), 1257–1272.

Fan, J. and Y. K. Truong (1993). Nonparametric regression with errors in variables. *Ann. Statist. 21*(4), 1900–1925.

Fan, J. and Q. Yao (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika 85*(3), 645–660.

Fan, Y. and Y. Liu (1997). A note on asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A 59*(1), 138–141.

Figueroa-López, J. E. (2009). Nonparametric estimation of Lévy models based on discrete–sampling. In *Optimality*, Volume 57 of *IMS Lecture Notes Monogr. Ser.*, pp. 117–146. Beachwood, OH: Inst. Math. Statist.

Figueroa-López, J. E. (2011). Sieve–based confidence intervals and bands for Lévy densities. *Bernoulli 17*(2), 643–670.

Figueroa-López, J. E. and C. Houdré (2006). Risk bounds for the non–parametric estimation of Lévy processes. In *High dimensional probability*, Volume 51 of *IMS Lecture Notes Monogr. Ser.*, pp. 96–116. Beachwood, OH: Inst. Math. Statist.

Giné, E. and R. Nickl (2008). Uniform central limit theorems for kernel density estimators. *Probab. Theory Related Fields 141*(3-4), 333–387.

Giné, E. and R. Nickl (2009). Uniform limit theorems for wavelet density estimators. *Ann. Probab. 37*(4), 1605–1646.

Girardi, M. and L. Weis (2003). Operator-valued Fourier multiplier theorems on Besov spaces. *Math. Nachr. 251*, 34–51.

Goldenshluger, A. and S. V. Pereverzev (2003). On adaptive inverse estimation of linear functionals in Hilbert scales. *Bernoulli 9*(5), 783–807.

Grafakos, L. (2004). *Classical and Modern Fourier Analysis.* Pearson Education, New Jersey.

Grama, I. and M. Nussbaum (2002). Asymptotic equivalence for nonparametric regression. *Math. Methods Statist. 11*(1), 1–36.

Gugushvili, S. (2009). Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *J. Nonparametr. Stat. 21*(3), 321–343.

Hall, P. and S. N. Lahiri (2008). Estimation of distributions, moments and quantiles in deconvolution problems. *Ann. Statist. 36*(5), 2110–2134.

Härdle, W. (1990). *Applied nonparametric regression.* Cambridge University Press.

Jacod, J. and A. N. Shiryaev (2003). *Limit theorems for stochastic processes* (Second ed.), Volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Berlin: Springer–Verlag.

Kahane, J.-P. (1985). *Some random series of functions* (Second ed.), Volume 5 of *Cambridge Studies in Advanced Mathematics.* Cambridge: Cambridge University Press.

Kappus, J. J. (2012). *Nonparametric adaptive estimation for discretely observed Lévy processes.* Ph. D. thesis, Humboldt–Universität zu Berlin. Available at http://edoc.hu-berlin.de/dissertationen/kappus-julia-johanna-2012-09-28/PDF/kappus.pdf.

Klenke, A. (2007). *Probability Theory: A Comprehensive Course (Universitext)* (1st ed.). Springer, London.

Le Cam, L. and G. L. Yang (2000). *Asymptotics in statistics* (Second ed.). Springer Series in Statistics. New York: Springer–Verlag. Some basic concepts.

Lounici, K. and R. Nickl (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Statist. 39*(1), 201–231.

Low, M. G. (1997). On nonparametric confidence intervals. *Ann. Statist. 25*(6), 2547–2554.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *J. Finan. Econ. 3*(1-2), 125 – 144.

Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Statist. 7*(4), 307–330.

Neumann, M. H. and M. Reiß (2009). Nonparametric estimation for Lévy processes from low–frequency observations. *Bernoulli 15*(1), 223–248.

Nickl, R. (2006). Empirical and Gaussian processes on Besov classes. In *High dimensional probability*, Volume 51 of *IMS Lecture Notes Monogr. Ser.*, pp. 185–195.

Nickl, R. and B. M. Pötscher (2007). Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov– and Sobolev–type. *J. Theoret. Probab. 20*(2), 177–199.

Nickl, R. and M. Reiß (2012). A Donsker theorem for Lévy measures. *J. Funct. Anal. 263*(10), 3306–3332.

Qui, B. H. (1981). Bernstein's theorem and translation invariant operators. *Hiroshima Math. J. 11*(1), 81–96.

Radulović, D. and M. Wegkamp (2000). Weak convergence of smoothed empirical processes: beyond Donsker classes. In *High dimensional probability, II*, Volume 47 of *Progr. Probab.*, pp. 89–105. Birkhäuser, Boston.

Sato, K.-i. (1999). *Lévy Processes and Infinitely Divisible Distributions*, Volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge: Cambridge University Press. Translated from the 1990 Japanese original, revised by the author.

Schmidt–Hieber, J., A. Munk, and L. Dümbgen (2012). Multiscale Methods for Shape Constraints in Deconvolution: Confidence Statements for Qualitative Features. *Ann. Statist.*, to appear. arXiv 1107.1404.

Söhl, J. (2010). Polar sets for anisotropic Gaussian random fields. *Statist. Probab. Lett. 80*(9-10), 840 – 847.

Söhl, J. (2012). Confidence sets in nonparametric calibration of exponential Lévy models. arXiv:1202.6611.

Söhl, J. and M. Trabs (2012a). A uniform central limit theorem and efficiency for deconvolution estimators. *Electron. J. Stat. 6*, 2486–2518.

Söhl, J. and M. Trabs (2012b). Option calibration of exponential Lévy models: Confidence intervals and empirical results. *J. Comput. Finance*, to appear. arXiv:1202.5983.

Tankov, P. (2011). Pricing and hedging in exponential Lévy models: review of recent results. In *Paris–Princeton Lectures on Mathematical Finance 2010*, Volume 2003 of *Lecture Notes in Math.*, pp. 319–359. Berlin: Springer.

Trabs, M. (2012). Calibration of selfdecomposable Lévy models. *Bernoulli*, to appear. arXiv:1111.1067.

Triebel, H. (2010). *Theory of Function Spaces*. Birkhäuser Verlag, Basel. Reprint of the 1983 Edition.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer, New York.

van Es, A. J. and H.-W. Uh (2004). Asymptotic normality of nonparametric kernel type deconvolution density estimators: crossing the Cauchy boundary. *J. Nonparametr. Stat. 16*(1-2), 261–277.

van Es, B. and H.-W. Uh (2005). Asymptotic normality of kernel–type deconvolution estimators. *Scand. J. Statist. 32*(3), 467–483.

Weber, M. (1983). Polar sets of some Gaussian processes. In *Probability in Banach spaces, IV (Oberwolfach, 1982)*, Volume 990 of *Lecture Notes in Math.*, pp. 204–214. Berlin: Springer.

Xiao, Y. (2009). Sample path properties of anisotrophic Gaussian random fields. Khoshnevisan, Davar (ed.) et al., A minicourse on stochastic partial differential equations. Papers based on the presentations at the minicourse, Salt Lake City, UT, USA, May 8–19, 2006. Berlin: Springer. Lecture Notes in Mathematics 1962, 145-212.

Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist. 18*(2), 806–831.

List of Figures

List of Tables

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 12.12.2012 Jakob Söhl